

Result No.	Score	Query Match	Length	DB	ID	Description
1	335	100.0	58	23	AAE16198	B. bronchiseptica
2	335	100.0	922	13	AAE25578	Bordetella parapertussis
3	335	100.0	922	23	AAE16185	Bordetella parapertussis
4	320	95.5	922	12	AAE14321	Pertactin antigen
5	314	93.7	60	23	AAE16195	B. bronchiseptica
6	303	90.4	56	23	AAE16197	B. bronchiseptica
7	284	84.8	52	23	AAE16194	B. bronchiseptica
8	284	84.8	911	12	AAE14320	Pertactin antigen
9	284	84.8	911	13	AAE26503	prn proteins, B. pertussis
10	284	84.8	911	23	AAE16183	Bordetella bronchiseptica

RESULT 14

Query Match 79.1%; Score 265; DB 23; Length 54;
Best Local Similarity 84.5%; Pred. No. 4.6e-13;
Matches 49; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

XX	B. bronchiseptica strain II-7 pertactin outer membrane protein region II.
DE	Pertactin; PN; outer membrane protein; vaccine; Bordetella infection;
XX	therapy; antibiotic; antibacterial; region II.
KW	
KW	
XX	Bordetella bronchiseptica.
OS	

AAE16193
ID AAE16193 standard; peptide; 49 AA.

DT 26-MAR-2002 (first entry)

DT	26-MAR-2002 (first entry)
XX	
DE	B. bronchiseptica strain II-1 pertactin outer membrane protein region II.

XX
XX PA (INSP) INST PASTEUR.
XX
XX
XX

Bordetella bronchiseptica.

1
2
3
4
5
6
7
8
9
10
11
12

29-NOV-2001.

23-MAY-2001; 2001WO-EP06457.

25-MAY-2000: 2000HS-206060P

(INSP) INST PASTEUR.

Guiso-maclouf N, Boursaux-eude C:

WPI; 2002-097639/13.

to identify antibodies to bacteria in materials such as human or other animal tissue and human or other animal cells, as well as biological fluids, such as human or other animal body fluids, including human sera, and to determine the concentration of Ab in those materials. Thus the antigens can be used for qualitative or quantitative determination of Bordetella in a material. The present sequence is B. bronchiseptica pertactin outer membrane protein region II.

Claim 26; Fig 1c; 47pp; English.

```

cc      peractin Outer membrane protein region II.
xx
SQ      Sequence 48 AA;
        Query Match          76.7%; Score 257; DB 23; Length 48;
        Best Local Similarity 82.8%; Pred. No. 1.6e-12;
        Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1
Qy      1 GAKAPPAKPAPQGPPQPQQPQPQPQPQPQPQPQPQPQPQPAPQPAPGRLSAA 58
         |||||
Db      1 GAKAPPAKPAPDQC-----PQPQPQPQPQPQPQPQPQPQPAPQPAPGRLSAA 48
         |||||
RESULT 15

```

Sequence 49 AA:

AAE16200
ID AAE16200 standard: peptide: 52 AA.

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	322	100.0	56	9	US-09-855-754-18		Sequence 18, Appl
2	303	94.1	58	9	US-09-855-754-19		Sequence 19, Appl
3	303	94.1	922	9	US-09-855-754-6		Sequence 6, Appl
4	294.5	91.5	59	9	US-09-855-754-16		Sequence 16, Appl
5	285	88.5	52	9	US-09-855-754-15		Sequence 15, Appl
6	285	88.5	4	9	US-09-855-754-4		Sequence 4, Appl
7	282	87.6	52	9	US-09-855-754-17		Sequence 17, Appl
8	264.5	82.1	49	9	US-09-855-754-14		Sequence 14, Appl
9	258	80.1	48	9	US-09-855-754-20		Sequence 20, Appl
10	257	79.8	52	9	US-09-855-754-21		Sequence 21, Appl
11	256	79.5	54	9	US-09-855-754-22		Sequence 22, Appl
12	210	65.2	42	9	US-09-855-754-23		Sequence 23, Appl
13	210	65.2	910	9	US-09-855-754-5		Sequence 5, Appl
14	189.5	58.9	39	9	US-09-855-754-24		Sequence 24, Appl
15	133	41.3	538	9	US-09-976-740-43		Sequence 43, Appl
16	133	41.3	538	12	US-10-023-529-43		Sequence 43, Appl
17	133	41.3	538	12	US-10-023-523-43		Sequence 43, Appl
18	132.5	41.1	550	9	US-09-976-740-47		Sequence 47, Appl
19	132.5	41.1	550	12	US-10-023-529-47		Sequence 47, Appl

RESULT 2
US-09-855-754-19
; Sequence 19, Application US/09855754
; Publication No. US2002019237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAUX-EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CO

```
; CURRENT APPLICATION NUMBER: US/09/855,754  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 59  
; TYPE: PRT  
; ORGANISM: Bordetella bronchiseptica  
  
US-09-855-754-16
```

Query Match 91.5%; Score 294.5; DB 9; Length 59;
Best Local Similarity 91.7%; Pred. No. 8.9e-14;

Matches 55; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

```
QY      1 GAKAPPAPKPA-----PQGPGGCGGGPPQPQQPPGPQPQRPEAPAPOPPAGRELSAA 56  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB       1 GAKAPPAPKPAQPQGPQGPQGPQGPQGPQPQQPPQPQR--PEAPAPOPPAGRELSAA 59  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 5

US-09-855-754-15

Sequence 15, Application US/09855754

Publication No. US20020192237A1

GENERAL INFORMATION:

APPLICANT: BOURSAXUX-EUDE, CAROLINE

TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

TITLE OF INVENTION: BORDETELLA PARAPTERTISSIS, AND BORDETELLA

TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

FILE REFERENCE: 03495-0206-00000

CURRENT APPLICATION NUMBER: US/09/855,754

CURRENT FILING DATE: 2001-09-10

PRIOR APPLICATION NUMBER: 60/206,969

PRIOR FILING DATE: 2000-05-25

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 15

LENGTH: 52

TYPE: PRT

ORGANISM: Bordetella bronchiseptica

US-09-855-754-15

Query Match 88.5%; Score 285; DB 9; Length 52;
Best Local Similarity 92.9%; Pred. No. 3.4e-13;

Matches 52; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

```
QY      1 GAKAPPAPKPAPQGPQGPQGPQGPQGPQPQQPPQPQRPEAPAPOPPAGRELSAA 56  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB       1 GAKAPPAPKPA----PQGGPGGPGPPQPQQPPQPQRPEAPAPOPPAGRELSAA 52  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 6

US-09-855-754-4

Sequence 4, Application US/09855754

Publication No. US20020192237A1

GENERAL INFORMATION:

APPLICANT: BOURSAXUX-EUDE, CAROLINE

TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

TITLE OF INVENTION: BORDETELLA PARAPTERTISSIS, AND BORDETELLA

TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

FILE REFERENCE: 03495-0206-00000

CURRENT APPLICATION NUMBER: US/09/855,754

CURRENT FILING DATE: 2001-09-10

PRIOR APPLICATION NUMBER: 60/206,969

PRIOR FILING DATE: 2000-05-25

NUMBER OF SEQ ID NOS: 24

RESULT 15
US-09-976-740-43
; Sequence 43, Application US/09976740
; Publication No. US20020194633A1

```

; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-43

```

```

Query Match      41.3%; Score 133; DB 9; Length 538;
Best Local Similarity 47.4%; Pred. No. 0.026;
Matches 27; Conservative 2; Mismatches 18; Indels 10; Gaps 2;

QY      2 AKAPPAPKPAQPGQPGQPGP-----POPPQPPQP-ORQPEAPAPQPP 48
      |  |||| |  |  ||: |  |  |  |||| |  |  |  ||||
Db      178 AAPPPAAPAVAPPAGPRAPPVAAAREPPLPPPPQPPAPPPQPPQPPQPP 234

```

Search completed: May 7, 2003, 17:30:52
Job time : 16.7905 secs


```
Query Match      88.5%; Score 285; DB 2; Length 911;
Best Local Similarity 92.9%; Pred. No. 6.6e-12;
Matches 52; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
```


Query Match 44.7%; Score 144; DB 2; Length 534;
Best Local Similarity 54.9%; Pred. NO. 0.0051;
Matches 28; Conservative 3; Mismatches 14; Indels

nuclear protein EBNA2 - human herpesvirus 4
C;Species: human herpesvirus 4, Epstein-Barr virus

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.43782 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-18

Perfect score: 322

Sequence: 1 GAKAPAPKAPQPGQPGP.....QRQPEAPAPQPPAGRELSAA 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	303	94.1	922	1 PERT_BORPA	P24328 bordetella
2	285	88.5	911	1 PERT_BORBR	Q03035 bordetella
3	210	65.2	910	1 PERT_BORPE	P14283 bordetella
4	147.5	45.8	449	1 APG_BRANA	P40603 brassica na
5	146.5	45.5	3164	1 TEGU_HSV11	P10220 herpes simp
6	146	45.3	141	1 YPRO_OWEFU	P21460 owenia fusi
7	146	45.3	485	1 SSGP_VOLCA	P21997 volvox cart
8	142.5	44.3	534	1 APG_ARATH	P40602 arabidopsis
9	141.5	43.9	487	1 EBN2_EBV	P12978 Epstein-Bar
10	140	43.5	2004	1 MOZ_HUMAN	Q92794 homo sapien
11	138.5	43.0	431	1 ACRO_RABIT	P48038 oryctolagus
12	136	42.2	555	1 GPL_CHLRE	Q9f946 chlamydomon
13	133.5	41.5	875	1 Y066_NPPOP	Q83949 orgyia pseu
14	132.5	41.1	426	1 EXLP_TOBAC	Q03211 nicotiana t
15	131.5	40.8	105	1 COLU_HSV7	P25050 herpesvirus
16	130.5	40.5	565	1 MOTB_MOUSE	O70324 mus musculu
17	126.5	39.3	102	1 COLL_HSVSC	P22576 herpesvirus
18	126.5	39.3	129	1 PARB_TRYBB	P09791 trypanosoma
19	126.5	39.3	261	1 PRP2_MOUSE	P05142 mus musculu
20	126.5	39.3	296	1 PRP3_MOUSE	P05143 mus musculu
21	126.5	39.3	1059	1 CAPU_DROME	Q24120 drosophila
22	126	39.1	440	1 CAPT_MOUSE	Q64467 mus musculu
23	125.5	39.0	1206	1 FMI4_MOUSE	Q05859 mus musculu
24	125.5	39.0	1468	1 FMI1_MOUSE	Q05860 mus musculu
25	124.5	38.7	143	1 PARL_TRYBB	P08469 trypanosoma
26	124.5	38.7	145	1 PARC_TRYBB	Q06884 trypanosoma
27	124	38.5	1790	1 SEPA_EMENI	P78621 emeritella
28	123.5	38.4	339	1 CSP_PLABE	P06915 plasmodium
29	123.5	38.4	347	1 CSP_PLABA	P23093 plasmodium
30	123.5	38.4	3110	1 HD_RAT	P51111 rattus norv
31	123	38.2	296	1	P04726 tritium ae
32	122	37.9	520	1 WASE_MOUSE	P70315 mus musculu
33	122	37.9	979	1 RFX1_HUMAN	P22670 homo sapien

RESULT 1

PERT_BORPA

ID PERT_BORPA

AC P24328

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Pertactin precursor (Outer membrane protein P.70) (P.95).

GN PRN.

OS Bordetella parapertussis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

OC Bordetella.

OX NCBI_TaxID=519;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CN2591;

RX MEDLINE=91251771; PubMed=2041476;

RA Li L.J., Dougan G., Novotny P., Charles I.G.;

RT "P.70 pertactin, an outer-membrane protein from Bordetella

parapertussis: cloning, nucleotide sequence and surface expression in

Escherichia coli."

RT Mol. Microbiol. 5:409-417(1991).

CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS

MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN

BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.

CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.

CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)

CONCENTRATIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

or send an email to license@isb-sib.ch).

CC EMBL: X54547; CAA38419.1; -

CC EMBL: A26124; CAA01786.1; -

CC EMBL: A19182; CAA01454.1; -

CC PIR: S15204; S15204.

CC PIR: S14659; S14659.

CC InterPro: IPR004899; Pertactin_sup.

CC InterPro: IPR003992; pertactin.

CC InterPro: IPR003991; pertactin_vir.

CC Pfam: PF03212; Pertactin; 1.

CC PRINTS; PR01482; PERTACTIN.

CC PRINTS; PR01484; PERTACTIN.FAMILY.

CC Outer membrane; Signal; Virulence; Repeat.

CC SIGNAL 1 34 POTENTIAL.

CC CHAIN 35 922 P.95.

CC CHAIN 35 647 PERTACTIN (P.70).

CC PROPEP 648 922 POTENTIAL.

CC SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN

ALIGNMENTS

34	121.5	37.7	367	1	CSP_PLAYO	P06914 plasmodium
35	121.5	37.7	502	1	WASP_HUMAN	P42768 homo sapien
36	121.5	37.7	3144	1	HD_HUMAN	P42858 homo sapien
37	120.5	37.4	415	1	ACRO_PIG	P08001 sus scrofa
38	120.5	37.4	443	1	HXA3_MOUSE	P02831 mus musculu
39	120.5	37.4	1048	1	SRA4_RAT	O03627 rattus norv
40	120	37.3	1362	1	BRD4_HUMAN	O60885 homo sapien
41	119	37.0	267	1	EXTN_MAIZE	P14918 zea mays (m
42	118.5	36.8	172	1	PRP2_RAT	P10164 rattus norv
43	118.5	36.8	642	1	MB11_ARATH	O9sav1 arabidopsis
44	118.5	36.8	1139	1	KPCL_TRIRE	Q99014 trichoderma
45	118	36.6	206	1	PRP3_RAT	P04474 rattus norv


```

DR Pfam: PF00657; Lipase_GDSL; 1.
DR PROSITE: PS01096; LIPASE_GDSL_SER; 1.
FT NON_TER 1
FT ACT_SITE 132 BY SIMILARITY.
FT ACT_SITE 428 POTENTIAL.
SQ SEQUENCE 449 AA; 48779 MW; 9EFB6A3AB28EA15 CRC64;

Query Match 45.88; Score 147.5; DB 1; Length 449;
Best Local Similarity 44.08; Pred. No. 0.0053;
Matches 29; Conservative 3; Mismatches 16; Indels 17; Gaps 2;

QY 3 KAPPAPKPPAQPGQPGQPGQPGQPPQPQ-----PPQPPQROPE---APAP 45
Db 3 KPQKPPPKPPQKPPPPAPTSPCPQPPKPPKPPAPTSPCPQPPKPPKPPPPAGP 62
QY 46 QPPAG 50
Db 63 SPKPG 67

RESULT 5
TEGU_HSV11
ID TEGU_HSV11 STANDARD; PRT: 3164 AA.
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Varion protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
CC 1- FUNCTION: TEGUMENT PROTEIN.
CC 1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC HSV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14112; CAA32311.1; -.
CC PIR: I30085; WMBEH6.
CC InterPro: IP005210; Herpes_UL36.
CC Pfam: PF03385; Herpes_UL36; 1.
CC Repeat.
KW DOMAIN
FT FT
SQ SEQUENCE 3164 AA; 335857 MW; CC5D31FF4F9PE3F4 CRC64;

Query Match 45.58; Score 146.5; DB 1; Length 3164;
Best Local Similarity 67.38; Pred. No. 0.025;
Matches 33; Conservative 2; Mismatches 9; Indels 5; Gaps 5;

QY 6 PAPKPPKPPQPGQPGQPGQ-PQP-PQP-PQP-PQRPAPAPAPQPPAG 50
Db 2935 PQPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 2982

RESULT 6
YPRO_OWEFU
ID YPRO_OWEFU STANDARD; PRT: 141 AA.

```

P21260; P21261;
AC 01-MAY-1991 (Rel. 18, Created)
DT
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical proline-rich protein (Fragment).
OS Oenota fusiformis.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palmeta; Canalipalpata;
OC Sabellidae; Owenillidae; Owenia.
OX NCBI_TaxID=6347;

[1]
SEQUENCE FROM N.A.
MEDLINE=90147742; PubMed=2105723;
RX Bakalara N., Collet J., Planellas R., Thouveny Y., Fontes M.;
RA "Presence in invertebrate genomes of sequences characterized by the
RT repetition of the triplet CCGPurine";
RL Biochem Biophys Res Commun. 166:66-73(1990).
DR FIR; A34043; A34043.
DR PIR; B34043; B34043.
KW Hypothetical protein; DNA-binding.
FT NON_TER 1
FT DOMAIN 9 58 POLY-PRO.
FT DNA_BIND 98 116 H-T-H MOTIF (POTENTIAL).
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15745 MW; B294E884D152BDB9 CRC64;

Query Match 45.3%; Score 146; DB 1; Length 141;
Best Local Similarity 53.2%; Pred.No. 0.0028;
Matches 25; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 5 PPAPKPAQPQGPGGPPQQFPQPPQPPQPPQPPQPPQPPQPPAQPAGR 51
II IIII II II II II II II II II II II II II II II II II II
DB 13 PPPPPPppPR 59

RESULT 7
SSGP_VOLICA STANDARD; PRT; 485 AA.

ID SSGP_VOLICA
AC P21997;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Sulfated surface glycoprotein 185 (SSG 185).
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvaceae; Volvox.
OX NCBI_TaxID=3067;

[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
STRAIN=f Nagariensis / HK10;
RX MEDLINE=90094551; PubMed=2689458;
RA "The extracellular matrix of Volvox carteri; molecular structure of
RT the cellular compartment".
RL J. Cell Biol. 109:3493-3501(1989).

CC -|- FUNCTION: THE EXTRACELLULAR MATRIX (ECM) OF VOLVOX CONTAINS
CC INSOLUBLE FIBROUS LAYERS THAT SURROUND INDIVIDUAL CELLS AT A
CC DISTANCE TO FORM CONTIGUOUS CELLULAR COMPARTMENTS. SSG 185 IS THE
CC MONOMERIC PRECURSOR OF THIS SUBSTRUCTURE (C3Z STRUCTURE). THE
CC COVALENT CROSS-LINKS ARE FORMED BETWEEN THE SACCHARIDE CHAINS
CC RATHER THAN BETWEEN THE POLYPEPTIDE CHAINS.
CC -|- PTM: A CENTRAL DOMAIN OF 77 AA CONSISTS ALMOST EXCLUSIVELY OF
CC HYDROXYPROLINE RESIDUES.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensel@lsb-sib.ch).

CC EMBL; X51616; CAA35953.1; --
DR PIR; A33647; A33647.

DR Pfam; PF00628; PHD; 2.
 DR Pfam; PF01853; MOZ_SAS; 1.
 DR SMART; SM00526; H15; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; ZF_PHD.1; 1.
 DR PROSITE; PS50016; ZF_PHD.2; 2.
 KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KW Nuclear protein.
 FT ZN_FING 206 265 PHD-TYPE 1.
 FT ZN_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT ZN_FING 538 560 C2HC-TYPE.
 FT DOMAIN 788 801 POLY-GLU.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547
 FT SITE
 SQ SEQUENCE 2004 AA; 225054 MW; 9FFBBAC3792854BA CRC64;
 Query Match 43.5%; Score 140; DB 1; Length 2004;
 Best Local Similarity 60.0%; Pred. No. 0.043;
 Matches 27; Conservative 1; Mismatches 15; Indels 2; Gaps 2;
 QY 5 PPAPKPAPQPGPGQPPQPPQ-PQPPQPPQPPQPPQPPQPP 48
 DB 1652 PPPPPQPPQ-PPPPQPPQPPPPPPQPPQPPQPPQPPPP 1695
 RESULT 11
 ACRO_RABIT STANDARD; PRT; 431 AA.
 ID ACRO_RABIT STANDARD; PRT; 431 AA.
 AC P48038;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Acrosin precursor (EC 3.4.21.10).
 GE ACN.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-New Zealand white; TISSUE-Testis;
 EX MEDLINE=9436861; PubMed=8086468;
 RA Richardson R.T., O'Rand M.G.
 RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
 preproacrosin-related cDNA.";
 RL Biochim. Biophys. Acta 1219:215-218(1994).
 CC -! FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
 CC IT IS A SERINE PROTEASE OF TRYPsin-LIKE CLEAVAGE SPECIFICITY, IT
 CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
 CC ACROSOME.
 CC -! CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
 CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
 CC -! SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

```

CC -1- SUBUNIT: Associates with GP2 and GP3.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF309494; AAC45420.1; -
CC EMBL: M58496; AAC69706.1; ALT_SEQ.
CC GlycerSutedB; Q9FPQ6; -
CC InterPro: IPR002965; P-rich_extensin.
CC InterPro: IPR003882; Pistil_extensin.
CC PRINTS: PR01217; PRICHEXTENSIN.
CC PRINTS: PR01218; PSTLEXTENSIN.
CC KW Glycoprotein; Repeat; Signal.
CC FT SIGNAL 1 29 POTENTIAL.
CC FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
CC FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.
CC FT DOMAIN 259 279 POLY-PRO.
CC FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 42.2%; Score 136; DB 1; Length 555;
Best Local Similarity 53.1%; Pred. No. 0.029;
Matches 26; Conservative 2; Mismatches 17; Indels 4; Gaps 2;

QY 4 APAPK-PAPQPGPQPGPQPPQP---PQPQPPQPPQPPQPPQPP 48
| | | | | | | | | | | | | | | | | | | | | |
Db 255 APPSPKPPAPPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 303

RESULT 13
Y056_NPVOP STANDARD; PRT; 875 AA.
ID Y056_NPVOP
AC Q83949; Q83964; O10323;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 98.6 kDa protein (ORE71).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OPMPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
RN [2]
RP SEQUENCE OF 1-807 FROM N.A.
RX MEDLINE=96201426; PubMed=8609478;
RA Ahrens C.H., Rohrmann G.F.;
RT "The DNA polymerase and helicase genes of a baculovirus of Orgyia
RT pseudotsugata.";
RL J. Gen. Virol. 77:825-837(1996).
RN [3]
RP SEQUENCE OF 806-875 FROM N.A.
RX MEDLINE=95343549; PubMed=7618274;
RA Ahrens C.H., Carlson C., Rohrmann G.F.;
RT "Identification, sequence, and transcriptional analysis of lef-3, a
RT gene essential for Orgyia pseudotsugata baculovirus DNA
RT replication.";
RL Virology 210:372-382(1995).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV AND LDMNPV.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75930; AAC59070.1; -
CC EMBL: U39145; AAB04047.1; -
CC EMBL: D45397; BAA08236.1; -
CC HSSP: P01100; IPOS.
CC KW Hypothetical protein.
CC FT DOMAIN 86 91 POLY-PRO.
CC SQ SEQUENCE 875 AA; 98603 MW; F03DB1B430D33C42 CRC64;

Query Match 41.5%; Score 133.5; DB 1; Length 875;
Best Local Similarity 56.8%; Pred. No. 0.057;
Matches 25; Conservative 2; Mismatches 16; Indels 1; Gaps 1;

QY 5 PPAPKPAFQPGPQPGPQPPQPPQPPQPPQPPQPPQPPQPPQPP 47
| | | | | | | | | | | | | | | | | | | | | |
Db 103 PPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 146

RESULT 14
EXLP_TOBAC STANDARD; PRT; 426 AA.
ID EXLP_TOBAC
AC Q03211;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pistil-specific extensin-like protein precursor (PELP).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Petit Havana; TISSUE=Pistil;
RC MEDLINE=93005740; PubMed=1392607;
RA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;
RT "Developmental expression of tobacco pistil-specific genes encoding
RT novel extensin-like proteins.";
RL Plant Cell 4:1041-1051(1992).
CC -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER
CC PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING
CC FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE
CC AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER
CC POLLINATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z14019; CAA78397.1; -
CC PIR: JQ1696; JQ1696.
CC InterPro: IPR000419; Pollen_Ole_e_I.
CC Pfam: PF01190; Pollen_Ole_e_I; 1.
CC KW Structural protein; Signal; Repeat; Glycoprotein.
CC FT SIGNAL 1 23
CC FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.
CC FT DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).
CC FT REPEAT 69 73 1.
CC FT REPEAT 76 80 2.
CC FT REPEAT 83 87 3.

```

Query Match 41.1%; Score 132.5; DB 1; Length 426;
Best Local Similarity 50.0%; Pred. No. 0.039;
Matches 26; Conservative 3; Mismatches 18; Indels 5

```

2 AKAPPAPKPAQPGQPGP- - - QPQPQPQPQPQPQPQPQPQPQPQPQPQP 48
|| || || || || || || || || || || || || || || || || || || || ||
165 AKQPQPQPAPAKQSPFPFPPPPVPVKAPSFAKQPPFPFPPPPVPVKAFSPATQPP 216

```

SULT 15
LL_HSV57
COLL_HSV57
STANDARD; PRT; 105 AA.
P25050;
01-MAY-1992 (Rel. 22, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
Collagen-like protein
Herpesvirus saimiri (strain 484-77).
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Rhadinovirus.
NCBI_TaxID=10382;
[1]
SEQUENCE FROM N.A.
MEDLINE=90279084; PubMed=2161952;
Gack P., Whitaker S.A., Medveczky M.M., Medveczky P.G.;
"Expression of collagenlike sequences by a tumor virus
saimiri";
J. Virol. 64:3509-3515(1990).

MEDLINE=92046383; PubMed=1658399;
 Geck P., Whitaker S.A., Medveczky M.M., Medveczky P.G.;
 "Expression of collagenlike sequences by a tumor virus, herpesvirus
 saimiri";
 J. Virol. 65:7084-7084(1991).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@sib-sib.ch).

PIR; A36770; CGBEHS.
 InterPro: IPR000087; Collagen.
 Pfam: PF01391; Collagen; 1.
 Collagen; Repeat.
 DOMAIN 15 70 COLLAGEN-LIKE.
 SEQUENCE 105 AA; 10260 MW; EF3DF0FE0FB446F0 CRC64;

Search completed: May 7, 2003, 16:48:10
 Elapsed time : 6.43782 secs

Result No.	Score	%			DB	ID	Description
		Match	Length				
1	322	100.0	111	2	Q9K5G9	Q9K5g9 bordetella	
2	322	100.0	132	2	Q9KJX9	Q9Kjx9 bordetella	
3	322	100.0	252	2	Q9ALQ2	Q9alq2 bordetella	
4	310	96.3	115	2	Q9K5H3	Q9K5h3 bordetella	
5	303	94.1	111	2	Q9K5G7	Q9K5g7 bordetella	
6	291.5	90.5	109	2	Q9K5H1	Q9K5h1 bordetella	
7	291.5	90.5	215	2	Q9ALQ4	Q9alq4 bordetella	
8	285	88.5	107	2	Q9K5H5	Q9K5h5 bordetella	
9	285	88.5	200	2	Q9ALP9	Q9alp9 bordetella	
10	285	88.5	911	2	Q9L4E2	Q9l4e2 bordetella	
11	264.5	82.1	104	2	Q9K5H9	Q9K5h9 bordetella	
12	264.5	82.1	105	2	Q9K5H7	Q9K5h7 bordetella	
13	264.5	82.1	115	2	Q9KJY1	Q9Kjy1 bordetella	
14	264.5	82.1	115	2	Q9KJY0	Q9Kjy0 bordetella	
15	264.5	82.1	198	2	Q9AHP0	Q9ahp0 bordetella	
16	258	80.1	195	2	Q9K5G5	Q9K5g5 bordetella	

RESULT	15
Q9AHP0	
ID	Q9AHP0 PRELIMINARY; PRN; 198 AA.
AC	Q9AHP0;
DT	01-JUN-2001 (TREMBlrel. 17, Created)
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Pertactin (Fragment).
GN	PRN.

Search completed: May 7, 2003, 16:53:14
Job time : 28.6201 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 25.46 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-14
Perfect score: 278
Sequence: 1 GAKAPPAPKPPGPPGPP.....*...QRQPEAPAPPPAGRELSAA 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Watch 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	278	100.0	49	23	AAE16193
2	266.5	95.9	52	23	AAE16194
3	266.5	95.9	911	12	AAE14320
4	266.5	95.9	911	13	AAE26503
5	266.5	95.9	911	23	AAE16183
6	266	95.7	53	23	AAE16196
7	264.5	95.1	56	23	AAE16197
8	263.5	94.8	58	23	AAE16198
9	263.5	94.8	922	12	AAE14321
10	263.5	94.8	922	13	AAE25578

11	263.5	94.8	922	23	AAE16185	Bordetella paraper
12	262.5	94.4	60	23	AAE16195	B. bronchiseptica
13	261.5	94.1	48	23	AAE16199	B. bronchiseptica
14	238.5	85.8	52	23	AAE16200	B. bronchiseptica
15	237.5	85.4	54	23	AAE16201	B. bronchiseptica
16	213.5	76.8	42	23	AAE16202	B. bronchiseptica
17	213.5	76.8	910	23	AAE16184	Bordetella pertuss
18	213.5	76.8	910	23	AAE17146	Bordetella pertuss
19	193	69.4	39	23	AAE16203	B. bronchiseptica
20	156.5	56.3	31	12	AAE11737	Peptide 683 derive
21	155.5	55.9	31	12	AAE11739	Peptide BBO5 deriv
22	128.5	46.2	24	17	AAE87649	Synthetic Bordetel
23	128	46.0	900	21	AAE42321	Human ORFX ORF2085
24	128	46.0	1217	22	ABG09876	Novel human diagno
25	128	46.0	1239	22	ABG09877	Novel human diagno
26	127	45.7	446	22	ABG00663	Drosophila melanog
27	126.5	45.5	406	22	ABG27250	Novel human diagno
28	125	45.0	76	23	ABG78537	Ser-Pro-Pro-Pro ge
29	123	44.2	85	22	AAO02036	Human polypeptide
30	123	44.2	88	22	AAO02124	Human polypeptide
31	122	43.9	80	22	AAO04412	Human polypeptide
32	122	43.9	99	22	AAO02076	Human polypeptide
33	121.5	43.7	44	22	AAO04929	Human polypeptide
34	121	43.5	85	22	AAO04573	Human polypeptide
35	119.5	43.0	325	22	ABG21919	Novel human diagno
36	119.5	43.0	356	22	ABG04360	Novel human diagno
37	118.5	42.6	350	23	ABG93838	Herbicidally activ
38	116.5	41.9	538	22	AAE82806	Rabbit low density
39	116.5	41.9	550	22	AAE82807	Rabbit low density
40	115.5	41.5	439	13	AAE28150	Sugar beet chitina
41	115	41.4	250	21	AAE67470	Np70 protein carbo
42	115	41.4	641	21	AAE82327	Human Npw38Bp1 tra
43	115	41.4	641	21	AAE67469	Np70 protein seque
44	115	41.4	641	22	AAE72165	Human RNA metaboli
45	115	41.4	641	22	AAE47514	NpwBP. Homo sapie

ALIGNMENTS

RESULT 1
AAE16193
ID AAE16193 standard; peptide; 49 AA.
XX
AC AAE16193;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-1 pertactin outer membrane protein region II.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
XX (INSP) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

The present invention relates to Bordetella bronchiseptica pertactin (outer membrane protein) or their fragments. Pertactin (PRN) is used as vaccine. Pertactin antibody is useful for treating Bordetella infections and used to detect Bordetella antigens in biological preparations or in purifying corresponding proteins, glycoproteins or their mixtures when used in affinity chromatographic columns. Pertactin is useful as antigens to identify antibodies to Bordetella in materials such as human or other animal tissue and human or other animal cells, as well as biological fluids, such as human or other animal body fluids, including human sera, and to determine the concentration of Ab in those materials. Thus the antigens can be used for qualitative or quantitative determination of


```

RESULT 6
AAE16196
ID AAE16196 standard; peptide; 53 AA.
AC AAE16196;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-4 pertactin outer membrane protein region II.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-4 pertactin outer membrane protein region II.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP ) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
PS Claim 26; Fig 1c; 47pp; English.
XX
CC The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region II.
XX
SQ Sequence 53 AA;
Query Match 95.78; Score 266; DB 23; Length 53;
Best Local Similarity 92.5; Pred. No. 3.1e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 1 GAKAPPAPKPA-----PQPGPQPGPQPPQPPQPPQPPQPPQPPAGRELSAA 49
Db 1 GAKAPPAPKPAQPGPQPGPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 53
RESULT 7
AAE16197
ID AAE16197 standard; peptide; 56 AA.
AC AAE16197;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-5 pertactin outer membrane protein region II.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX

```

```

XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP ) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
PS Claim 26; Fig 1c; 47pp; English.
XX
CC The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region II.
XX
SQ Sequence 56 AA;
Query Match 95.18; Score 264.5; DB 23; Length 56;
Best Local Similarity 87.58; Pred. No. 4.2e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 7; Gaps 1;
QY 1 GAKAPPAPKPAQPGPQPG-----PQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 49
Db 1 GAKAPPAPKPAQPGPQPGPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 56
RESULT 8
AAE16198
ID AAE16198 standard; peptide; 58 AA.
AC AAE16198;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-6 pertactin outer membrane protein region II.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP ) INST PASTEUR.
XX

```


44-38861-1000

ID AAE16201 standard; peptide; 54 AA.

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	266.5	95.9	911	4	US-08-460-269C-4	Sequence 4, Appli
2	263.5	94.8	922	4	US-08-460-269C-6	Sequence 6, Appli
3	213.5	76.8	910	4	US-08-460-269C-2	Sequence 2, Appli
4	131.5	47.3	33	1	US-08-237-716-11	Sequence 11, Appl
5	128.5	46.2	24	4	US-08-750-624-11	Sequence 11, Appl
6	121.5	43.7	23	4	US-08-460-269C-8	Sequence 8, Appli
7	114	41.0	3119	1	US-08-246-982A-16	Sequence 16, Appl
8	114	41.0	3119	1	US-08-453-285-16	Sequence 16, Appl
9	112	40.3	75	4	US-09-314-268-149	Sequence 149, App
10	111.5	40.1	334	6	5202236-3	Patent No. 5202236
11	110	39.6	380	2	US-09-026-587-4	Sequence 4, Appli
12	110	39.6	380	2	US-09-227-420-4	Sequence 4, Appli
13	108	38.8	478	3	US-08-153-888-2	Sequence 2, Appli
14	107.5	38.7	513	4	US-09-041-886-28	Sequence 28, Appl
15	107.5	38.7	530	4	US-09-041-886-29	Sequence 29, Appl
16	107.5	38.7	552	4	US-09-041-886-30	Sequence 30, Appl
17	107.5	38.7	589	4	US-09-041-886-31	Sequence 31, Appl
18	107.5	38.7	1274	4	US-09-095-443-2	Sequence 2, Appli
19	107.5	38.7	3144	1	US-08-246-982A-6	Sequence 6, Appli
20	107.5	38.7	3144	1	US-08-453-285-6	Sequence 6, Appli
21	107.5	38.7	3144	2	US-08-457-273B-42	Sequence 42, Appl
22	107.5	38.7	3144	3	US-08-556-419-21	Sequence 21, Appl
23	107.5	38.7	3144	4	US-09-041-886-15	Sequence 15, Appl
24	106	38.1	3118	2	US-08-457-273B-8	Sequence 8, Appli
25	105.5	37.9	331	6	5202236-37	Patent No. 5202236
26	105.5	37.9	402	6	5352575-4	Patent No. 5352575
27	104.5	37.6	366	6	5470718-4	Patent No. 5470718

```

Query Match          95.98;  Score 266.5;  DB 4;  Length 911;
Best Local Similarity 94.28;  Pred. No. 1.6e-16;
Matches 49;  Conservative 0;  Mismatches 10;  Indels 3;  Gaps 1;

1  GAKAPPKPKAPQPGPGG---PQPPQPPQPPQPPQPPQPPAGRELSAA 49
559 GAKAPPKPKAPQPGPGPGPPPPPPPPQPPQPPQPPQPPQPPAGRELSAA 610

```


NAME/KEY: Region
LOCATION: 1..2
OTHER INFORMATION: /label= LTB sequence
FEATURE:
NAME/KEY: Region
LOCATION: 4..7
OTHER INFORMATION: /label= hinge region
FEATURE:
NAME/KEY: Region
LOCATION: 10..31
OTHER INFORMATION: /label= p69 BB05 epitope of Bordetella
OTHER INFORMATION: pertussis sequence
US-08-237-716-11

Query Match 47.3%; Score 131.5; DB 1; Length 33;
Best Local Similarity 74.2%; Pred. NO. 6e-06;
Matches 23; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

Qy 10 PAPGPGPQPQPQPQPQPQPQPQPQP 40
Db 5 PGPETAPQPGPQPQPQPQPQPQPQP 32

RESULT 5
US-08-750-624-11
Sequence 11, Application US/08750624
Patent No. 6290971
GENERAL INFORMATION:
APPLICANT: Kandil, Ali
APPLICANT: James, Olive A
APPLICANT: Chong, Pele
APPLICANT: Klein, Michael H
TITLE OF INVENTION: ADJUVANT COMPOSITIONS COMPRISING A
TITLE OF INVENTION: MINERAL SALT AND ANOTHER IMMUNOSTIMULATING COMPOUND
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,624
FILING DATE: 26-FEB-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-660
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-750-624-11
Query Match 46.2%; Score 128.5; DB 4; Length 24;
Best Local Similarity 88.5%; Pred. NO. 8.2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 16 POPGPGPQPQPQPQPQPQPQPQP 41
Db 1 POPGPGPQPQPQPQPQPQPQPQP 23

RESULT 6
US-08-460-269C-8
Sequence 8, Application US/08460269C
Patent No. 6197548
GENERAL INFORMATION:
APPLICANT: CLARE, JEFFREY J.
ROMANOS, MICHAEL A.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
YEAST
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
STREET: 2200 Clarendon Blvd., Suite 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6333
TELEFAX: (703) 243-6410
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-460-269C-8

Query Match 43.7%; Score 121.5; DB 4; Length 23;
Best Local Similarity 88.0%; Pred. NO. 3.3e-05;
Matches 22; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 16 POPGPGPQPQPQPQPQPQPQPQP 40
Db 2 POPGPGPQPQPQPQPQPQPQP 23

RESULT 7
US-08-246-982A-16
Sequence 16, Application US/08246982A
Patent No. 5686288
GENERAL INFORMATION:
APPLICANT: MacDonald, Marcy E.
APPLICANT: Ambrose, Christine M.
APPLICANT: Puyao, Mabel P.
APPLICANT: Gusella, James F.
TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Query Match 41.0%; Score 114; DB 1; Length 3119;

ESULT 11
S-09-026-587-4
Sequence 4, Application US/09026587
Patent No. 5912128
GENERAL INFORMATION:

APPLICANT: Rabizadeh, Sharroo
TITLE OF INVENTION: Proapoptotic Peptides, Dependence

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	278	100.0	49	9	US-09-855-754-14	Sequence 14, Appl
2	266.5	95.9	52	9	US-09-855-754-15	Sequence 15, Appl
3	266.5	95.9	911	9	US-09-855-754-4	Sequence 4, Appl
4	264.5	95.1	56	9	US-09-855-754-18	Sequence 18, Appl
5	263.5	94.8	58	9	US-09-855-754-19	Sequence 19, Appl
6	263.5	94.8	922	9	US-09-855-754-6	Sequence 6, Appl
7	261.5	94.1	48	9	US-09-855-754-20	Sequence 20, Appl
8	250.5	90.1	52	9	US-09-855-754-17	Sequence 17, Appl
9	247	88.8	59	9	US-09-855-754-16	Sequence 16, Appl
10	238.5	85.8	52	9	US-09-855-754-21	Sequence 21, Appl
11	237.5	85.4	54	9	US-09-855-754-22	Sequence 22, Appl
12	213.5	76.8	42	9	US-09-855-754-23	Sequence 23, Appl
13	213.5	76.8	910	9	US-09-855-754-5	Sequence 5, Appl
14	193	69.4	39	9	US-09-855-754-24	Sequence 24, Appl
15	117	42.1	827	9	US-10-171-384-3	Sequence 3, Appl
16	116.5	41.9	538	9	US-09-976-740-43	Sequence 43, Appl
17	116.5	41.9	538	12	US-10-023-529-43	Sequence 43, Appl
18	116.5	41.9	538	12	US-10-023-523-43	Sequence 43, Appl
19	116.5	41.9	550	9	US-09-976-740-47	Sequence 47, Appl

Search completed: May 7, 2003, 17:30:50
Job time : 15.6917 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	266.5	95.9		911	2	A47675	68K outer membrane
2	263.5	94.8		922	2	S15204	pactactin - Bordet
3	213.5	76.8		910	2	A32560	outer membrane pro
4	132.4	47.5		449	2	S16748	proline-rich prote
5	131	47.1		430	2	JC2301	hypothetical 47.8K
6	131	47.1		430	2	T10798	hypothetical 47.8K
7	125	45.0		599	2	S42442	phosphorin-S - VO
8	124.5	44.8		929	2	C96232	nuclear protein EB
9	124.5	44.8		929	2	C96232	hypothetical prote
10	124.5	44.8		1684	2	T02632	hypothetical prote
11	122	44.6		485	2	A33647	sulfated surface g
12	122	43.9		141	2	A34043	hypothetical proli
13	121.5	43.7		544	2	T17547	proline-rich prote
14	121.5	43.7		209	2	A48232	cysteine-rich exte
15	121.5	43.5		1952	2	T48814	hypothetical prote
16	120.5	43.3		370	2	WBEH6	UL36 protein - num
17	120	43.2		446	2	H83619	hypothetical prote
18	119.5	43.0		534	2	S21961	hydroxyproline-ric
19	118	42.4		1137	2	A86335	proline-rich prote
20	117.5	42.3		440	2	IA9681	T20H2.9 protein -
21	117.5	42.3		1494	2	T14355	glyceraldehyde-3-p
22	117	42.1		383	2	A43295	protein-tyrosine-p
23	117	42.1		875	2	T10340	hypothetical prote
24	116.5	41.9		412	2	T17531	hypothetical prote
25	116	41.7		426	2	JQ1696	proline-rich prote
26	115.5	41.5		439	2	S51939	pitill extensin-II
27	115	41.4		165	2	T24470	chitinase (EC 3.2.
28	115	41.4		196	2	B48232	hypothetical prote
29	115	41.4		415	1	A34170	cysteine-rich exte
30	115	41.4		415	1	A34170	acrosin (EC 3.4.21

QY 1 GA KAPPA KPA PQ

QY 1 GAKAPPAKPAQPGPG-----PQPQPQPQPQPQPQPQPAGRELSAA 49

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 4.75809 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-14
Perfect score: 278
Sequence: 1 GAKAPPAPKPAQPGQPGP.....ORQPEAPAPQPPAGRELSAA 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266.5	95.9	911	1 PERT_BORBR	Q03035 bordetella
2	263.5	94.8	922	1 PERT_BORPA	P24328 bordetella
3	213.5	76.8	910	1 PERT_BORPE	P14283 bordetella
4	132	47.5	449	1 APC_BRANA	P40603 brassica na
5	125	45.0	487	1 EBN2_EBV	P12978 Epstein-Bar
6	124	44.6	485	1 SSGP_VOLCA	P21997 volvox cart
7	122	43.9	141	1 YPRO_OWEPU	P21260 owenla fusi
8	121	43.5	2004	1 MOZ_HUMAN	Q92794 homo sapien
9	121	43.5	3164	1 TEGU_HSV11	P10220 herpes simp
10	118	42.4	534	1 APC_ARATH	P40602 arabidopsis
11	117.5	42.3	440	1 G3PT_MOUSE	Q64467 mus musculu
12	117	42.1	875	1 Y066_NPVOP	Q83949 orgyia pseu
13	116	41.7	426	1 EXLP_TOBAC	Q03211 nicotiana t
14	116	41.7	443	1 HXA3_MOUSE	P02831 mus musculu
15	116	41.7	555	1 GPI_CHLRE	Q91pq6 chlamydomon
16	115	41.4	415	1 ACRO_PIG	P08001 sus scrofa
17	112.5	40.5	1362	1 BRD4_HUMAN	Q60885 homo sapien
18	111.5	40.1	431	1 ACRO_RABIT	P06318 oryctolagus
19	111	39.9	339	1 CSP_PLABE	P06915 plasmodium
20	111	39.9	347	1 CSP_PLABA	P23093 plasmodium
21	111	39.9	565	1 MOT8_MOUSE	Q70324 mus musculu
22	111	39.9	1139	1 KFC1_TRIRE	Q99014 trichoderma
23	111	39.9	1790	1 SEPA_EMENI	P78621 emericella
24	110	39.6	380	1 VASP_HUMAN	P50552 homo sapien
25	110	39.6	3110	1 HD_RAT	P51111 rattus norv
26	109.5	39.4	757	1 CLKF_HUMAN	Q14003 homo sapien
27	108	38.8	367	1 CSP_PLAYO	P06914 plasmodium
28	108	38.8	808	1 Y066_NPVAC	P41467 autographa
29	107.5	38.7	501	1 WASL_RAT	Q08816 rattus norv
30	107.5	38.7	505	1 WASL_BOVIN	Q95107 bos taurus
31	107.5	38.7	505	1 WASL_HUMAN	Q00401 homo sapien
32	107.5	38.7	3144	1 HD_HUMAN	P42858 homo sapien
33	107	38.5	268	1 NOZ0_MEDTR	P93329 medicago tr

RESULT 1

ID	PERT_BORBR	STANDARD;	PRT;	911 AA.
AC	Q03035;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.68) (P.94).			
GN	PRN.			
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=518;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.			
RC	STRAIN=CN7531;			
RX	MEDLINE=92407514; PubMed=1527510;			
RA	Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;			
RT	"Cloning, nucleotide sequence and heterologous expression of the			
RT	protective outer-membrane protein P.68 pertactin from Bordetella			
RT	bronchiseptica."			
RL	J. Gen. Microbiol. 138:1697-1705(1992).			
CC	-!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS			
CC	MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN			
CC	BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-!- SUBUNIT: MONOMER.			
CC	-!- SUBCELLULAR LOCATION: Outer membrane.			
CC	-!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)			
CC	CONCENTRATIONS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X54815; CAA38584.1; -			
DR	EMBL; A19180; CAA01453.1; -			
DR	PIR; A47675; A47675.			
DR	InterPro; IPR004899; Pertactin_sup.			
DR	InterPro; IPR003992; pertactin.			
DR	InterPro; IPR003991; pertactin_vir.			
DR	Pfam; PF03212; Pertactin_1.			
DR	PRINTS; PR01432; PERTACTIN.			
DR	PRINTS; PR01434; PRVACTNFAMLY.			
DR	Outer membrane; Signal; Virulence; Repeat.			
KW	SIGNAL	1	34	
FT	CHAIN	35	911	P.94.
FT	CHAIN	35	712	PERTACTIN (P.68).
FT	PROPEP	713	911	POTENTIAL.
FT	SITE	250	262	CELL ATTACHMENT SITE (INVOLVED IN
FT				ADHESION TO VARIOUS EUKARYOTIC CELL
FT				LINES).

ALIGNMENTS

34	106.5	38.3	267	1	EXTN_MAIZE	P14918 zea mays (m
35	106.5	38.3	514	1	MEFD_MOUSE	Q63943 mus musculu
36	106.5	38.3	2715	1	TRX2_HUMAN	Q9umh6 homo sapien
37	106	38.1	232	1	ACRL_HUMAN	P58840 homo sapien
38	106	38.1	261	1	PRP2_MOUSE	P05142 mus musculu
39	106	38.1	296	1	PRP3_MOUSE	P05143 mus musculu
40	106	38.1	642	1	MB11_ARATH	Q9sav1 arabidopsis
41	106	38.1	979	1	RFX1_HUMAN	P22670 homo sapien
42	106	38.1	1206	1	FMN1_MOUSE	Q05859 mus musculu
43	106	38.1	1468	1	FMN1_MOUSE	Q05860 mus musculu
44	106	38.1	3119	1	HD_MOUSE	P42859 mus musculu
45	105.5	37.9	402	1	VGLD_PPRVI	P07645 pseudorabie

```

FT FT ADHESION TO VARIOUS EUKARYOTIC CELL
FT FT LINES).
FT FT 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT FT 1.
FT FT 2.
FT FT 3.
FT FT 4 (APPROXIMATE).
FT FT 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
FT FT SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
SQ
Query Match 94.8%; Score 263.5; DB 1; Length 922;
Best Local Similarity 84.5%; Pred. No. 2.4e-10;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPAPKAPQPGQPG-----PQPPQPPQPPQPPAPQPPAGRELSAA 49
DB 564 GAKAPAPKAPQPGQPGQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 621

RESULT 3
PERT_BORPE
ID ID PERT_BORPE STANDARD; PRT; 910 AA.
AC AC P14283;
DT DT 01-JAN-1990 (Rel. 13, Created)
DT DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE Pertactin precursor (Outer membrane protein P.69) (P.93).
GN GN OR OMP69A.
OS Bordetella pertussis.
OS Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OC OX NCBI_TaxID=520;
RN RN [1]
RN RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC RC STRAIN=CN2992;
RX RX MEDLINE=89264462; PubMed=2542937;
RA RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT RT "Molecular cloning and characterization of protective outer membrane
RT RT protein P.69 from Bordetella pertussis.";
RN RN Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RN RN [2]
RN RN REVISIONS TO 264 AND 332.
RX RX MEDLINE=92407534; PubMed=1527510;
RX RX Li J.-L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT RT "Cloning, nucleotide sequence and heterologous expression of the
RT RT protective outer-membrane protein P.68 pertactin from Bordetella
RT RT bronchiseptica.";
RN RN J. Gen. Microbiol. 138:1697-1705(1992).
RN RN [3]
RN RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX RX MEDLINE=96196517; PubMed=8609998;
RA RA Emmsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RN RN Nature 381:90-92(1996).
CC CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC CC -1- SUBUNIT: MONOMER.
CC CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CC CONCENTRATION.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; J04560; AAA22980.1; ALT SEQ.

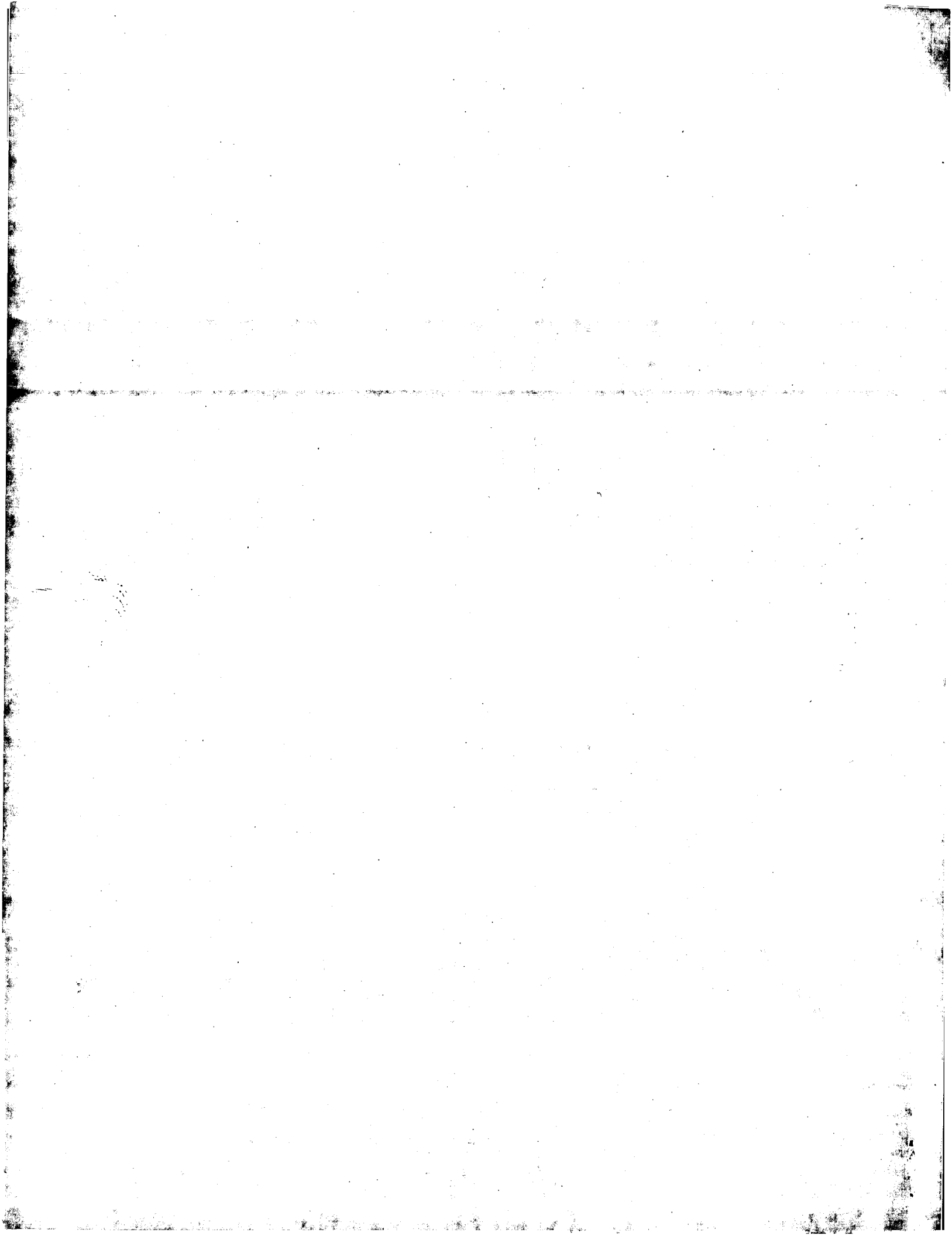
```


SQ SEQUENCE 200 AA; 20826 MW; 9F3AC6E4128942B6 CRC64;
 Query Match 95.9%; Score 266.5; DB 2; Length 200;
 Best Local Similarity 94.2%; Pred. No. 4.5e-16;
 Matches 49; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GAKAPPAKFAPOPGPPG---PQPPQPQQPEAPAPQPPAGRELSAA 49
 Db 127 GAKAPPAKFAPOPGPPGPPGPPQPPQPPQPPQPEAPAPQPPAGRELSAA 178
 |||||||
 RESULT 8
 Q9L4E2 PRELIMINARY; PRT; 911 AA.
 ID Q9L4E2
 AC Q9L4E2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Pertactin.
 GN PRN
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B14;
 RC King A.; Berbers G., Hoogerhout P., Olirschot van H.F., Knipping K.,
 RA Mooi F.R.;
 RT "Role of the polymorphic region of the B.pertussis protein pertactin
 RT in immunity: implications for the design of pertussis vaccines.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR ENBL: AJ245927; CAB82515.1;
 DR InterPro: IPR003992; pertactin.
 DR InterPro: IPR004899; Pertactin_vlr.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS; PRO1482; PERTACTIN.
 DR PRINTS; PRO1484; PRtACTINFAMLY.
 SQ SEQUENCE 911 AA; 94093 MW; DF531AE4383A32 CRC64;
 Query Match 95.9%; Score 266.5; DB 2; Length 911;
 Best Local Similarity 94.2%; Pred. No. 1.6e-15;
 Matches 49; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GAKAPPAKFAPOPGPPG---PQPPQPQQPEAPAPQPPAGRELSAA 49
 Db 559 GAKAPPAKFAPOPGPPGPPQPPQPPQPPQPEAPAPQPPAGRELSAA 610
 |||||||
 RESULT 9
 Q9K5H1 PRELIMINARY; PRT; 109 AA.
 ID Q9K5H1
 AC Q9K5H1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Pertactin (P.68) (Fragment).
 GN PRN
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CVGEO;
 RX MEDLINE=20359389; PubMed=10899896;
 RA Boursaux-Eude C., Guiso N.;
 RT "Polymorphism of Repeated Regions of Pertactin in Bordetella
 RT pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
 RL Infect. Immun. 68:4815-4817(2000).
 DR EMBL: AJ250085; CAB76439.1;
 DR InterPro: IPR004899; Pertact sup

[illegible]



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 27.0187 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754b-15
Perfect score: 297
Sequence: 1 GAKAPAPKAPQPPQPGP.....QRQPEAPAPQPPAGRELSAA 52

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	100.0	52	AAE16194	B. bronchiseptica
2	297	100.0	911	AAE14320	Pertactin antigen
3	297	100.0	911	AAE14320	prn proteins. Bor
4	297	100.0	911	AAE16183	Bordetella bronchi
5	286.5	96.5	53	AAE16196	B. bronchiseptica
6	285	96.0	56	AAE16197	B. bronchiseptica
7	284	95.6	58	AAE16198	B. bronchiseptica
8	284	95.6	922	AAE14321	Pertactin antigen
9	284	95.6	922	AAE14321	Bordetella parapap
10	284	95.6	922	AAE16185	Bordetella parapap

11	283	95.3	60	23	AAE16195	B. bronchiseptica
12	266.5	89.7	49	23	AAE16193	B. bronchiseptica
13	260	87.5	48	23	AAE16199	B. bronchiseptica
14	259	87.2	52	23	AAE16200	B. bronchiseptica
15	258	86.9	54	23	AAE16201	B. bronchiseptica
16	212	71.4	42	23	AAE16202	B. bronchiseptica
17	212	71.4	910	23	AAE16184	Bordetella pertuss
18	212	71.4	910	23	AAE17146	Bordetella pertuss
19	191.5	64.5	39	23	AAE16203	B. bronchiseptica
20	155	52.2	31	12	AAE11737	Peptide 683 derive
21	154	51.9	31	12	AAE11739	Peptide BR05 deriv
22	142.5	48.0	900	21	AB42321	Human ORFX ORF2085
23	142.5	48.0	1217	22	ABG09876	Novel human diago
24	142.5	48.0	1239	22	ABG09877	Novel human diago
25	135	45.5	80	22	AAO04412	Human polypeptide
26	135	45.5	99	22	AAO02076	Human polypeptide
27	134	45.1	85	22	AAO04573	Human polypeptide
28	134	45.1	406	22	ABG27250	Novel human diago
29	133	44.8	446	22	ABG70063	Drosophila melanog
30	131	44.1	44	22	AAO04929	Human polypeptide
31	130	43.8	88	22	AAO02124	Human polypeptide
32	130	43.8	439	13	AAE28150	Sugar beet chitina
33	129	43.4	325	22	ABG21919	Novel human diago
34	128.5	43.3	76	23	ABE78537	Ser-Pro-Pro-Pro ge
35	128.5	43.3	85	22	AAO02036	Human polypeptide
36	127.5	42.9	470	21	AAE49660	Arabidopsis thalia
37	127	42.8	24	17	AAE87649	Synthetic Bordetel
38	125.5	42.3	104	23	ABE78538	Ser-Pro-Pro-Pro-Pr
39	125	42.1	334	9	AAE82971	Bioadhesive precur
40	125	42.1	350	23	ABE93838	Herbicidally activ
41	123	41.4	538	22	ABE82806	Human low density
42	122.5	41.2	356	22	ABG04360	Novel human diago
43	122.5	41.2	1222	21	AAE36453	Arabidopsis thalia
44	122.5	41.2	1257	21	AAE36452	Arabidopsis thalia
45	122.5	41.2	1275	21	AAE36451	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE16194
ID AAE16194 standard; peptide; 52 AA.
XX
AC AAE16194;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-2 pertactin outer membrane protein region II.
XX
DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella Infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
XX WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EF06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
XX (INSP) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

PS Claim 1; Fig 1; 28pp; English.

Db

1	GAKAPPAKPAPQPGPQGPPGPPGPPOPPOROPPEAPOPPAGRELSAA	53
---	---	----

Sequence 48 AA;

ID	AAE16201 standard; peptide; 54 AA.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	297	100.0	911	4	US-08-460-269C-4	Sequence 4, Appli
2	284	95.6	922	4	US-08-460-269C-6	Sequence 6, Appli
3	212	71.4	910	4	US-08-460-269C-2	Sequence 2, Appli
4	130	43.8	33	1	US-08-237-716-11	Sequence 11, Appl
5	127	42.8	24	1	US-08-750-624-11	Sequence 11, Appl
6	125	42.1	334	6	52022336-3	Patent No. 52022336
7	120.5	40.6	3119	1	US-08-246-982A-16	Sequence 16, Appl
8	120.5	40.6	3119	1	US-08-453-265-16	Sequence 16, Appl
9	120	40.4	23	4	US-08-460-269C-8	Sequence 8, Appli
10	120	40.4	331	6	52022336-37	Patent No. 52022336
11	116	39.1	513	4	US-09-041-886-28	Sequence 28, Appl
12	116	39.1	530	4	US-09-041-886-29	Sequence 29, Appl
13	116	39.1	552	4	US-09-041-886-30	Sequence 30, Appl
14	116	39.1	589	4	US-09-041-886-31	Sequence 31, Appl
15	116	39.1	3144	1	US-08-246-982A-6	Sequence 6, Appli
16	116	39.1	3144	1	US-08-453-265-6	Sequence 6, Appli
17	116	39.1	3144	2	US-08-457-273B-42	Sequence 42, Appl
18	116	39.1	3144	3	US-08-558-419-21	Sequence 21, Appl
19	116	39.1	3144	4	US-09-041-886-15	Sequence 15, Appl
20	114.5	38.6	478	3	US-08-153-888-2	Sequence 2, Appli
21	113.5	38.2	106	4	US-09-314-268-134	Sequence 134, App
22	112.5	37.9	1274	4	US-09-095-443-2	Sequence 2, Appli
23	112.5	37.9	3118	2	US-08-457-273B-8	Sequence 8, Appli
24	111.5	37.5	641	4	US-08-961-083-160	Sequence 160, App
25	110.5	37.2	330	1	US-08-642-235-32	Sequence 32, Appl
26	110.5	37.2	408	1	US-07-609-716-65	Sequence 65, Appl
27	110.5	37.2	408	4	US-08-475-411A-65	Sequence 65, Appl

Db 1 POPGPOPPPOPP-----OPeAPAPopp 2

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,982A
FILING DATE: May 20, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldstein, Jorge, A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3880002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3119 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-246-982A-16

```


CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.

RESULT 14

```

1  REEL 15
2  US-08-246-982A-6
3  ; Sequence 6, Application US/08246982A
4  ; Patent No. 5686288
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: MacDonalad, Marcy E.
9  ; APPLICANT: Ambrose, Christine M.
10 ; APPLICANT: Duyao, Mabel P.
11 ; APPLICANT: Gusella, James F.
12 ; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
13 ; NUMBER OF SEQUENCES: 25
14 ; CORRESPONDENCE ADDRESS:
15 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
16 ; STREET: 1100 New York Avenue
17 ; CITY: Washington
18 ; STATE: D.C.
19 ; COUNTRY: U.S.A.
20 ; ZIP: 20005
21 ;
22 ; COMPUTER READABLE FORM:
23 ; MEDIUM TYPE: Floppy disk
24 ; COMPUTER: IBM PC compatible
25 ; OPERATING SYSTEM: PC-DOS/MS-DOS
26 ; SOFTWARE: Patent In Release #1.0, Version #1.25
27 ; CURRENT APPLICATION DATA:
28 ;

```

```

; APPLICATION NUMBER:  US/08/246,982A
; FILING DATE:  May 20, 1994
; CLASSIFICATION:  435
; ATTORNEY/AGENT INFORMATION:
; NAME:  Goldstein, Jorge, A.
; REGISTRATION NUMBER:  29,021
; REFERENCE/DOCKET NUMBER:  0609,3880002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (202) 371-2600
; TELEFAX:  (202) 371-2540
; INFORMATION FOR SEQ ID NO:  6:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  3144 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-08-246-982A-6

Query Match          39.1%; Score 116; DB 1; Length 3144;
Best Local Similarity 45.6%; Pred. No. 0.011;
Matches 26; Conservative 2; Mismatches 15; Indels 14; Gaps 2;

QY      9  KPAPQPGQPGPPPPQPPQPQ--PQRQPEAPAPQPPAG-----RELSA 51
      :| | | | | | | | | | | | | | | | | | | | |
Db      40  QPPPPPPPPPPQLPQPPPPQAQLLLPQPPPPPPPPPPGPAVAEPLHRPKKLSA 96
      :| | | | | | | | | | | | | | | | | | | | |

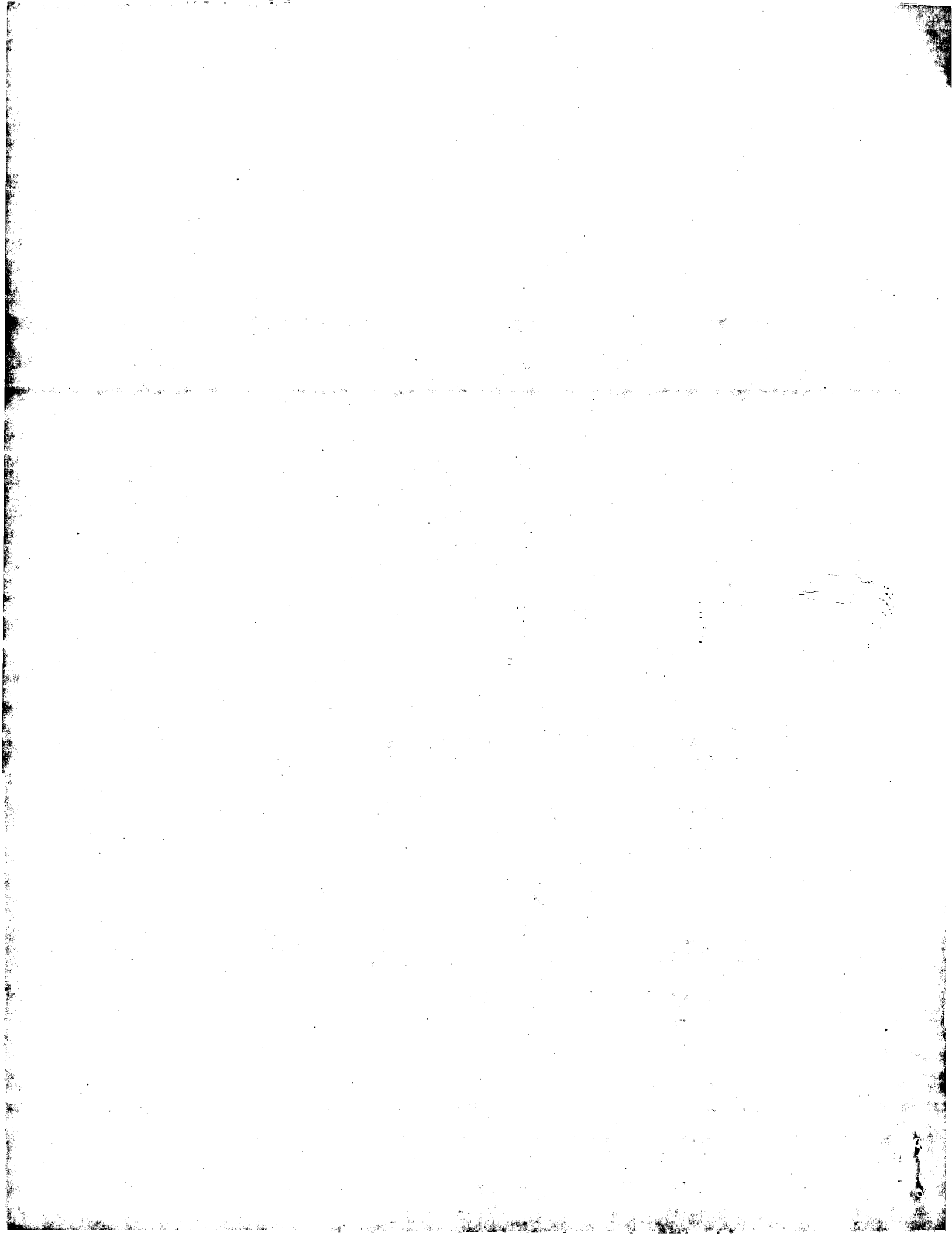
Search completed: May 7, 2003, 16:57:16
Job time : 23.6838 secs

```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	297	100.0	52	9	US-09-855-754-15	Sequence 15, Appl
2	297	100.0	911	9	US-09-855-754-4	Sequence 4, Appl
3	285	96.0	56	9	US-09-855-754-18	Sequence 18, Appl
4	284	95.6	58	9	US-09-855-754-19	Sequence 19, Appl
5	284	95.6	922	9	US-09-855-754-6	Sequence 6, Appl
6	271	91.2	52	9	US-09-855-754-17	Sequence 17, Appl
7	267.5	90.1	52	9	US-09-855-754-16	Sequence 16, Appl
8	266.5	89.7	49	9	US-09-855-754-14	Sequence 14, Appl
9	260	87.5	48	9	US-09-855-754-20	Sequence 20, Appl
10	259	87.2	52	9	US-09-855-754-21	Sequence 21, Appl
11	258	86.9	54	9	US-09-855-754-22	Sequence 22, Appl
12	212	71.4	42	9	US-09-855-754-23	Sequence 23, Appl
13	212	71.4	910	9	US-09-855-754-5	Sequence 5, Appl
14	191.5	64.5	39	9	US-09-855-754-24	Sequence 24, Appl
15	123	41.4	538	9	US-09-976-740-43	Sequence 43, Appl
16	123	41.4	538	12	US-10-023-529-43	Sequence 43, Appl
17	123	41.4	538	12	US-10-023-529-43	Sequence 43, Appl
18	121.5	40.9	550	9	US-09-976-740-47	Sequence 47, Appl
19	121.5	40.9	550	12	US-10-023-529-47	Sequence 47, Appl

US-09-855-754-4
; Sequence 4, Application US/09855754
; Publication No. US2020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAUX-EUDE, CAROLINE
; APPLICANT: GUIZO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES C

Search completed: May 7, 2003, 17:30:50
Job time : 15.5911 secs



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	297	100.0	911	2	A47675	68K outer membrane	
2	284	95.6	922	2	S15204	pertactin - Bordet	
3	212	71.4	930	2	A32560	outer membrane pro	
4	151.5	51.0	410	2	JC2301	hypothetical 47.8K	
5	141	47.1	559	2	T10798	phosphorin-S - Vo	
6	140.5	47.3	449	2	S15748	proline-rich prote	
7	137	46.1	485	2	A33647	sulfated surface g	
8	136	45.8	487	2	A42442	nuclear protein BB	
9	135	45.5	141	2	A34043	hypothetical proli	
10	135	45.5	209	2	A48232	cysteine-rich exte	
11	132	44.4	534	2	S21961	proline-rich prote	
12	132	44.4	1952	2	T48814	hypothetical prote	
13	130.5	43.9	1137	2	A85335	T20H2.9 protein -	
14	130	43.8	439	2	S51939	chitinase (EC 3.2.	
15	129.5	43.6	3164	1	WMBEH6	UL36 protein - hum	
16	129	43.4	383	2	AE2295	hypothetical prote	
17	129	43.4	544	2	T17547	proline-rich prote	
18	128.5	43.3	446	2	T07907	hydroxyproline-ric	
19	128	43.1	875	2	T10340	hypothetical prote	
20	128	43.1	929	2	C96623	hypothetical prote	
21	127	42.8	270	2	H83619	hypothetical prote	
22	127	42.8	760	2	T06291	extensin homolog T	
23	126	42.4	1494	2	T14355	protein-tyrosine-p	
24	125.5	42.3	196	2	B48232	cysteine-rich exte	
25	125.5	42.3	431	2	S47538	acrosin (EC 3.4.21	
26	123	41.4	1684	2	T03632	hypothetical prote	
27	122.5	41.2	494	2	B96534	hypothetical prote	
28	122.5	41.2	645	2	A71416	hypothetical prote	
29	122.5	41.2	1585	2	T31611	hypothetical prote	

Query Match	45.5%	Score 135;	DB 2;	Length 209;
Best Local Similarity	55.8%	Pred. No. 0.0064;		
Matches 24;	Conservative	4;	Mismatches 13;	Indels 2;
			Gaps	2;

Qy	5	P P A K P A F Q P G P G P Q P - P Q P P P Q P - P Q R Q P E A P A P Q P P A 45
Db	85	P P R P C F S P P P P P R P C S P S P P P P P Q P P R P R P S P P S P P P P A 127

RESULT 11
S21961
proline-rich protein APG - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997

C;Accession: S21961
R;Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, R.J.
submitted to the EMBL Data Library, August 1991.
A:Description: Cloning and characterization of a proline-rich gene expressed specifically
A:Reference number: S16748
A:Accession: S21961
A:Molecule type: DNA
A:Residues: 1-534 <ROB>
A:Cross-references: EMBL:X60377; NID:g22598; PID:g22599
C:Genetics:
A:Gene: APG
A:Introns: 256/1; 299/3; 387/3; 470/1

Query Match	44.4%	Score 132;	DB 2;	Length 534;
Best Local Similarity	52.0%	Pred. No. 0.021;		
Matches 26;	Conservative	3;	Mismatches 13;	Indels 8;
				Gaps 2;

[illegible]

```

RESULT 12
T48814
hypothetical protein 15E6.220 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48814
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.
submitted to the Protein Sequence Database, April 2000

```

Query Match 44.48; Score 132; DB 2; Length 1952;

```

RESULT 14
S51939
chitinase (EC 3.2.1.14) precursor - beet
C:Species: Beta vulgaris (beet)
C>Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 10-Jul-1998
C:Accession: S51939; S72315; S45025
R:Berglund, L.; Brunstedt, J.; Nielsen, K.K.; Chen, Z.; Mikkelsen, J.D.; Marcker, K.A.
Plant Mol. Biol. 27, 211-216, 1995
A>Title: A Proline-rich chitinase from Beta vulgaris.
A:Reference number: S51939; MOID:95170004; PMID:7865792
A:Accession: S51939
A:Molecule type: DNA
A:Residues: 1-439 <BER>
A:Cross-references: EMBL:X79301; NID:g488730; PID:g488731
A>Note: the authors translated the codon TGC for residue 416 as Gly
A:Accession: S72315
A:Molecule type: mRNA
A:Residues: 191-397 <BER2>
C:Genetics:
A:Introns: 248/1; 300/2
C:Superfamily: plant chitinase homology
C:Keywords: glycosidase; hydrolase
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-439/Product: chitinase #status predicted <MAT>
F:183-423/Domain: plant chitinase homology <PCH>

Query Match      43.8%; Score 130; DB 2; Length 439;
Best Local Similarity 48.8%; Pred. No. 0.024;
Matches 21; Conservative 2; Mismatches 20; Indels 0; Gaps 0;

QY      3 KAPPAPKPAQGPGQPQQPPQPPQPPQPPQPPQPEAPAQPPA 45

```


[illegible]

```

ZN_FING      538      560      C2HC-TYPE.
FT DOMAIN    788      801      POLY-GLU.
FT DOMAIN    989      995      POLY-GLU.
FT DOMAIN   1019     1026      POLY-ARG.
FT DOMAIN   1069     1078      POLY-GLU.
FT DOMAIN   1147     1150      POLY-LYS.
FT DOMAIN   1221     1242      GLU-RICH.
FT DOMAIN   1267     1302      GLU-RICH.
FT DOMAIN   1411     1414      POLY-GLU.
FT DOMAIN   1593     1597      POLY-SER.
FT DOMAIN   1643     1704      GLN/PRO-RICH.
FT DOMAIN   1897     1977      MET-RICH.
FT SITE      1546     1547      BREAKPOINT FOR TRANSLOCATION TO FORM
SQ SEQUENCE  2004 AA; 225054 MW; 9FFBBAC3792854BA CRC64;

Query Match          42.9%; Score 127.5; DB 1; Length 2004;
Best Local Similarity 55.6%; Pred. No. 0.16;
Matches 25; Conservative 1; Mismatches 14; Indels 5; Gaps 2;

QY 5 PPAPKP---APQGPQGPGQPQP-PQQPQPQRQPEAPAPQPP 44
    ||||| | | | | | | | | | | | | | | | | | |
Db 1651 PPPPPPQQPPPPQPPAPAPPPPPPPQQPQQPQPQPQPPPPPP 1695

RESULT 12
ACRO_RABIT STANDARD; PRT; 431 AA.
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
GN ACR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Testis;
RX MEDLINE=94368861; PubMed=8086468;
RA Richardson R.T., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
RT preproacrosin-related cDNA.";
RL Biochim. Biophys. Acta 1219:215-218(1994).
CC -1- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOEA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY. IT
CC IS SYNTHESIZED IN A ZYGOTE FORM, PROCROSIN AND STORED IN THE
CC ACROSOME.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -1- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U05204; AAA61630.1; -.
CC HSP; P00763; IDPO.
CC MEROPS; S01.223; -.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00689; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYP_SPC; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.

```

```

DR PROSITE; PS00134; TRYPSIN_HIS; 1.
KW PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16 BY SIMILARITY.
FT CHAIN 17 431 ACROSIN.
FT CHAIN 17 39 ACROSIN LIGHT CHAIN (BY SIMILARITY).
FT CHAIN 40 7 ACROSIN HEAVY CHAIN (BY SIMILARITY).
FT PROPEP 7 431 PRO-RICH.
FT DOMAIN 40 288 SERINE PROTEASE.
FT DISULFID 22 152 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 160 INTERCHAIN (BY SIMILARITY).
FT DISULFID 71 87 BY SIMILARITY.
FT DISULFID 175 244 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 234 264 BY SIMILARITY.
FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 431 AA; 46422 MW; 1C0154E0BC0C668 CRC64;

Query Match 42.3%; Score 125.5; DB 1; Length 431;
Best Local Similarity 43.6%; Pred. No. 0.071;
Matches 24; Conservative 0; Mismatches 20; Indels 11; Gaps 1;

OY 2 AKAPPAPKPAQPGQPQ-----PGQPQPQPQPQPQPQPQPQPQPQPQPQPQP 45
DB 325 ASGPPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHP 379

RESULT 13
GPI_CHLRE
ID GPI_CHLRE STANDARD; PRT; 555 AA.
AC Q9FPQ6; Q03927;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (hydroxyproline-rich
DE glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
CC Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -!- SUBUNIT: Associates with GP2 and GP3.
CC -!- PTM: N-glycosylated and O-glycosylated.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

DR EMBL; AF309494; AAG45420.1; -.
DR EMBL; M58496; AAG69706.1; ALT_SEQ.
DR GlycoSuiteDB; Q9FPQ6; -.
DR InterPro; IPR002965; P-rich_extensin.
DR PRINTS; IPR003882; Pistil_extensin.
DR PRINTS; PRO1217; PRICHEXTENSIN.
DR PRINTS; PRO1218; PSTLEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSXP REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 42.1%; Score 125; DB 1; Length 555;
Best Local Similarity 54.3%; Pred. No. 0.091;
Matches 25; Conservative 2; Mismatches 13; Indels 6; Gaps 3;

OY 4 APPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 45
DB 59 APPSPAP-PSGP-PSFAPPSPPSPAPPSPAPPSPAPPSPAPPSPAPPSPA 102

RESULT 14
HXA3_MOUSE
ID HXA3_MOUSE STANDARD; PRT; 443 AA.
AC P02831; Q61197;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A3 (Hox-1.5) (MO-10).
GN HOXA3 OR HOXA-3 OR HOX-1.5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Hofmann M., Boehm T.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 54-280 FROM N.A.
RC STRAIN=ICR Swiss;
RX MEDLINE=96323206; PubMed=8710855;
RA Tan D., Shao X., Pu L., Guo V., Nirenberg M.;
RT "Sequence and expression of the murine Hoxd-3 homeobox gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:8247-8252(1996).
RN [3]
RP SEQUENCE OF 185-258 FROM N.A.
RA Ruddle F.H., Hart C.P., McGinnis W.;
RT "Structural and functional aspects of the mammalian homeo-box
RT sequences.";
RN [4]
RP Trends Genet. 1:48-51(1985).
RN [5]
RP SEQUENCE OF 185-258 FROM N.A.
RX MEDLINE=85024859; PubMed=6091896;
RA McGinnis W., Hart C.P., Gehring W.J., Ruddle F.H.;
RT "Molecular cloning and chromosome mapping of a mouse DNA sequence
RT homologous to homeotic genes of Drosophila.";
RL Cell 38:675-680(1984).
RN [5]
RP DNA-BINDING.
RX MEDLINE=87092283; PubMed=2879282;
RA Fainsod A., Bogard L.D., Ruusala T., Lubin M., Crothers D.M.,
RA Ruddle F.H.;
RT "The homeo domain of a murine protein binds 5' to its own homeo box.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9532-9536(1986).
RN [6]
RP DEVELOPMENTAL STAGE.

```


45

```

DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS; PRO1217; PRICHEXTENS.
FT NON_TER 1
FT FT 122
SQ SEQUENCE 122 AA; 12395 MW; 8ED00966A40FF994 CRC64;

Query Match          96.08; Score 285; DB 2; Length 122;
Best Local Similarity 92.9%; Pred. No. 1.6e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPPAPKPA-----PPGPGPQGPPQPQQPPQPQQPEAPAPQPPAGRELSAA 52
   |||||      |||||||
DB 42 GAKAPPAPKPAQPGPGPQGPPQPQQPPQPQQPEAPAPQPPAGRELSAA 97
   |||||      |||||||

RESULT 8
Q9ALQ2 PRELIMINARY; PRT; 252 AA.
ID Q9ALO2 AC Q9ALO2:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pertactin (Fragment).
OS B.
GN Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19395;
RX MEDLINE=21117018; PubMed=11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella
   RT bronchiseptica Pertactin."
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL; JAY007265; AAC38441.1; -
DR InterPro: IPR003992; Pertactin.scp
DR InterPro: IPR004899; Pertactin.sup
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS; PRO1482; PERTACTIN.
DR PRINTS; PRO1217; PRICHEXTENS.
FT NON_TER 1
FT FT 252
SQ SEQUENCE 252 AA; 26107 MW; 368C142508D77057 CRC64;

Query Match          96.08; Score 285; DB 2; Length 252;
Best Local Similarity 92.9%; Pred. No. 3e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPPAPKPA-----PPGPGPQGPPQPQQPPQPQQPEAPAPQPPAGRELSAA 52
   |||||      |||||||
DB 170 GAKAPPAPKPAQPGPGPQGPPQPQQPPQPQQPEAPAPQPPAGRELSAA 225
   |||||      |||||||

RESULT 9
Q9K5G7 PRELIMINARY; PRT; 111 AA.
ID Q9K5G7 AC Q9K5G7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pertactin (P:68) (Fragment).
OS B.
GN Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEL:

```

OC	Bordetella.
OX	NCBI_TaxID=518;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=9.73H+;
RA	MEDLINE=20359389; PubMed=10899896;
RX	Boursaux-Eude C., Guiso N.;
RT	"Polymorphism of Repeated Regions of Pertactin in Bordetella
RL	pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RT	Infect. Immun. 68:4815-4817(2000).
RE	EMBL; AJ250077; CAB76431.1; -.
DR	InterPro; IPR004899; Pertact_sup.
DR	Pfam; PF03212; Pertactin; 1.
FT	NON_TER 1
FT	NON_TER 104
FT	NON_TER 104
SQ	SEQUENCE 104 AA; 10714 MW; 4C7248E86CFF1189 CRC64;
Query Match 89.7%; Score 266.5; DB 2; Length 104;	
Best Local Similarity 94.2%; Pred. No. 5.2e-16;	
Matches 49; Conservative 0; Mismatches 0; Indels 3; Gaps 1;	
QY	1 GAKAPPAPKPAPOGPGPGPPGPQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQPQQPAGRELSAA 52
Db	21 GAKAPPAPKPAPOGPGPG---PQQPQQPQQPQQPQQPQQPQQPQQPQQPAGRELSAA 69
RESULT 12	
O9K5H7	PRELIMINARY; PRT; 105 AA.
ID	O9K5H7
AC	O9K5H7;
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Pertactin (P.68) (Fragment).
GN	PN.
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC	Bordetella.
OX	NCBI_TaxID=518;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=LAPR.
RA	MEDLINE=20359389; PubMed=10899896;
RX	Boursaux-Eude C., Guiso N.;
RT	"Polymorphism of Repeated Regions of Pertactin in Bordetella
RL	pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RT	Infect. Immun. 68:4815-4817(2000).
RE	EMBL; AJ250079; CAB76433.1; -.
DR	InterPro; IPR004899; Pertact_sup.
DR	Pfam; PF03212; Pertactin; 1.
FT	NON_TER 1
FT	NON_TER 105
FT	NON_TER 105
SQ	SEQUENCE 105 AA; 10861 MW; 1F97255EB792F4B7 CRC64;
Query Match 89.7%; Score 266.5; DB 2; Length 105;	
Best Local Similarity 94.2%; Pred. No. 5.2e-16;	
Matches 49; Conservative 0; Mismatches 0; Indels 3; Gaps 1;	
QY	1 GAKAPPAPKPAPOGPGPGPPGPQPQQPQQPQQPQQPQQPQQPQQPQQPQQPAGRELSAA 52
Db	22 GAKAPPAPKPAPOGPGPG---PQQPQQPQQPQQPQQPQQPQQPQQPQQPAGRELSAA 70
RESULT 13	
O9KJY1	PRELIMINARY; PRT; 115 AA.
ID	O9KJY1
AC	O9KJY1;
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Pertactin (fragment).
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

Search completed: May 7, 2003, 16:53:13
Job time : 27.5758 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 30.6559 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-16
Perfect score: 342
Sequence: 1 GAKAPAPKAPQPPQPCP.....PORPEAPAPQPPAGRELSNA 59

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	331.5	96.9	60	23	AAE16195
2	298.5	87.3	58	23	AAE16198
3	298.5	87.3	922	13	AAE16185
4	298.5	87.3	922	13	AAE16185
5	294.5	86.1	56	23	AAE16197
6	283.5	82.9	922	12	AAE14321
7	274	80.1	53	23	AAE16196
8	267.5	78.2	52	23	AAE16194
9	267.5	78.2	911	12	AAE14320
10	267.5	78.2	911	13	AAE26503

11	267.5	78.2	911	23	AAE16183
12	249.5	73.0	54	23	AAE16201
13	247	72.2	49	23	AAE16193
14	242.5	70.9	52	23	AAE16200
15	240.5	70.3	48	23	AAE16199
16	203.5	59.5	42	23	AAE16202
17	203.5	59.5	910	23	AAE16184
18	203.5	59.5	910	23	AAE17146
19	183	53.5	39	23	AAE16203
20	158	46.2	80	22	AAO04412
21	158	46.2	99	22	AAO02076
22	157	45.9	85	22	AAO02036
23	157	45.9	85	22	AAO04573
24	154	45.0	76	23	ABB78537
25	154	45.0	900	21	AAE42321
26	154	45.0	1217	22	ABG09876
27	154	45.0	1239	22	ABG09877
28	152.5	44.6	446	22	ABB70063
29	151.5	44.3	31	12	AAE11737
30	148	43.3	406	22	ABG27250
31	146.5	42.8	88	22	AAO02124
32	146	42.7	325	22	ABG21919
33	146	42.7	439	13	AAE28150
34	145.5	42.5	31	12	AAE11739
35	145.5	42.5	129	18	AAW14573
36	145.5	42.5	190	18	AAW14569
37	145.5	42.5	350	23	ABB93838
38	144	42.1	183	18	AAW14570
39	143.5	42.0	206	18	AAW14574
40	142	41.5	104	23	ABB78538
41	142	41.5	329	23	AAE14305
42	142	41.5	594	22	ABB61362
43	141.5	41.4	358	21	AAE22949
44	140.5	41.1	334	9	AAE82971
45	140	40.9	541	19	AAW37148

ALIGNMENTS

RESULT 1
AAE16195
ID AAE16195 standard; peptide; 60 AA.
XX
AC AAE16195;
XX
XX 26-MAR-2002 (first entry)
DT
XX B. bronchiseptica strain II-3 pertactin outer membrane protein region II.
DE
DE Pertactin; PFN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
XX WO200190143-P2.
PN
XX 29-NOV-2001.
PD
XX 23-MAY-2001; 2001WO-EP06457.
PF
XX 25-MAY-2000; 2000US-206969P.
PR
XX (INSP) INST PASTEUR.
PA
XX Guiso-maclouf N, Boursaux-eude C;
PI
XX WPI; 2002-097639/13.
DR
XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

The present invention relates to Bordetella bronchiseptica pertactin (outer membrane protein) or their fragments. Pertactin (PRN) is used as vaccine. Pertactin antibody is useful for treating Bordetella infections and used to detect Bordetella antigens in biological preparations or in purifying corresponding proteins, glycoproteins or their mixtures when used in affinity chromatographic columns. Pertactin is useful as antigens to identify antibodies to Bordetella in materials such as human or other animal tissue and human or other animal cells, as well as biological fluids, such as human or other animal body fluids, including human sera, and to determine the concentration of Ab in those materials. Thus the antigens can be used for qualitative or quantitative determination of

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-32

Query Match 38.0%; Score 130; DB 1; Length 330;
Best Local Similarity 47.5%; Pred. No. 0.00014;
Matches 29; Conservative 0; Mismatches 24; Indels 8; Gaps 3;
QY 1 GAKAPPAPKAPGPGQPGP--QPGP-----OPG-PQPPQPPQPPQPPQPPQPPQPPA 52
DB 235 GPGPGPGPGAPGPGPGPGPGPGPGAGVGSFGAGPGPGPGPGPGPGPGPGPGPP 294
QY 53 G 53
DB 295 G 295

RESULT 12
US-07-609-716-65
Sequence 65, Application US/07609716
Patent No. 5514581
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.

REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-609-716-65

Query Match 38.0%; Score 130; DB 1; Length 408;
Best Local Similarity 47.5%; Pred. No. 0.00017;
Matches 29; Conservative 0; Mismatches 24; Indels 8; Gaps 3;
QY 1 GAKAPPAPKAPGPGQPGP--QPGP-----OPG-PQPPQPPQPPQPPQPPQPPQPPA 52
DB 313 GPGPGPGPGAPGPGPGPGPGPGPGAGVGSFGAGPGPGPGPGPGPGPGPGPP 372
QY 53 G 53
DB 373 G 373

RESULT 13
US-08-475-411A-65
Sequence 65, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,411A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 65:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-475-411A-65

Query Match          38.0%; Score 130; DB 4; Length 408;
Best Local Similarity 47.5%; Pred. No. 0.00017;
Matches 29; Conservative 0; Mismatches 24; Indels 8; Gaps 3;

QY 1 GAKAPPAPKPAQPGQPGP--QPGP-----QPG-PQPQPQPQPQPQPQPAPQPPA 52
Db 313 GPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 372
QY 53 G 53
Db 373 G 373

RESULT 14
US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
US-08-478-029A-65

Query Match          38.0%; Score 130; DB 4; Length 408;
Best Local Similarity 47.5%; Pred. No. 0.00017;
Matches 29; Conservative 0; Mismatches 24; Indels 8; Gaps 3;

QY 1 GAKAPPAPKPAQPGQPGP--QPGP-----QPG-PQPQPQPQPQPQPQPAPQPPA 52
Db 313 GPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 372
QY 53 G 53
Db 373 G 373

RESULT 15
US-08-961-083-160
; Sequence 160, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-160

Query Match          37.6%; Score 128.5; DB 4; Length 641;
Best Local Similarity 48.4%; Pred. No. 0.00034;
Matches 30; Conservative 3; Mismatches 20; Indels 9; Gaps 4;

QY 6 PAPKPAQ-PQPGQPGP-QPGQPGQPGPQPP-----PQPQPQPQPQPAPAPQPP 56
Db 348 PAPAPQPPQAPAPKPKPEQAPAPKPKPEQAPAPKPKPEQAPAPKPKPEQAP 407
QY 57 SA 58
Db 408 PA 409

Search completed: May 7, 2003, 16:57:17
Job time : 10.8405 secs
```


Nature 406: 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: H83619
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-270 <STO>
 A:Cross-references: GB:AE004458; GB:AE004091; NID:g9946031; PIDN:AAG03586.1; GSPDB:GN000001
 A:Experimental source: strain PA01

```

C:Genetics:
A:Gene: PA0197
C:Superfamily: tonB protein

Query Match      45.9%  Score 157;  DB 2;  Length 270;
Best Local Similarity 50.0%;  Pred. No. 0.00058;
Matches 29;  Conservative 4;  Mismatches 17;  Indels 8;  Gaps 1;

Qy      2  AKAPAPKAPAPQPGQPGQPGQPGQ-----PPQPGFPQPPQPPQPPQPPQPP  51
      || ||||| : : : ||| : : : ||||| |||||
Db      123  AVKPPPPKPKPKPKPKQPPRRPKAPKAVEPAPPQPPAAAPPAPAPAAAPLTP  180

RESULT 12
A48232
cysteine-rich extensin-like protein 1 precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)

```

C; Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 01-Dec-2000
C; Accession: A48232; PQ0475; S24617
R; Wu, H.; Zou, J.; May, B.; Gu, Q.; Cheung, A. Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 6829-6833, 1993
A; Title: A tobacco gene family for flower cell wall proteins with a proline-rich domain
A; Reference number: A48232; MUID:93342083; PMID:8341705
A; Accession: A48232
A; Status: preliminary

A:Residues: 1-209; <GB>
A:Cross-references: GB:113439; NID:g310922; PIDN:AAA34059.1; PID:g310923
R:de S Goldman, M.H.; Pezzotti, M.; Seurinck, J.; Mariani, C.
Plant Cell 4, 1041-1051, 1992
A:Title: Developmental expression of tobacco pistil-specific genes encoding novel ext
A:Reference number: PQ0474; MUID:93005740; PMID:1392607
A:Accession: PQ0475
A:Molecule type: mRNA
A:Residues: 39-209 <GOL>
A:Cross-references: EMBL:214020; NID:g19918; PID:g19919
A:Experimental source: stigma, style; strain Petit Havana SRI
C:Superfamily: glutelin
C:Keywords: cell wall; extracellular matrix; fertilization; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-209/Product: cysteine-rich extensin-like protein 1 #status experimental <MAT>
F:146/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 45.8%; Score 156.5; DB 2; Length 209;
Best Local Similarity 48.2%; Pred. No. 0.00051;
Query Match 45.8%; Score 156.5; DB 2; Length 209;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.72913 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-16

Perfect score: 342

Sequence: 1 GAKAPAPKAPQGPQGP.....PORPEAPAPQPPAGRELSAA 59

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	298.5	87.3	922	1 PERT_BORPA	P24328 bordetella
2	267.5	78.2	911	1 PERT_BORBR	Q03035 bordetella
3	203.5	59.5	910	1 PERT_BORPE	P14283 bordetella
4	161.5	47.2	449	1 APG_BRANA	P40603 brassica na
5	158	46.2	141	1 YPRO_OWEFU	P21260 owensia fusi
6	158	46.2	485	1 SSGP_VOLCA	P21997 volvox cart
7	158	46.2	3164	1 TEGU_HSVII	P10220 herpes simp
8	149.5	43.7	555	1 GPI_CHLRE	Q9fpq6 chlamydomon
9	148.5	43.4	487	1 EBN2_EBV	P12978 epstein-bar
10	148.5	43.4	534	1 APG_ARATH	P40602 arabidopsis
11	147	43.0	102	1 COLL_HSVSC	P22576 herpesvirus
12	145	42.4	261	1 PRP2_MOUSE	P05142 mus musculus
13	145	42.4	296	1 PRP3_MOUSE	P05143 mus musculus
14	144	42.1	431	1 ACRO_RABIT	P48038 oryctolagus
15	143	41.8	426	1 EXLP_TOBAC	Q03211 nicotiana t
16	141.5	41.4	875	1 Y056_NPVOP	Q83949 orgyia pseu
17	140.5	41.1	2004	1 MOZ_HUMAN	Q92794 homo sapien
18	139	40.6	129	1 PARB_TRYBB	P09791 trypanosoma
19	136	39.8	143	1 PARL_TRYBB	P08469 trypanosoma
20	136	39.8	145	1 PARC_TRYBB	Q06084 trypanosoma
21	136	39.8	502	1 WASP_HUMAN	P42768 homo sapien
22	135.5	39.6	520	1 WASP_MOUSE	P70315 mus musculus
23	134	39.2	565	1 MOT8_MOUSE	Q07324 mus musculus
24	133	38.9	339	1 CSP_PLABE	P06915 plasmodium
25	133	38.9	347	1 CSP_PLABA	P23093 plasmodium
26	133	38.9	1206	1 FM14_MOUSE	Q05859 mus musculus
27	133	38.9	1468	1 FMN1_MOUSE	Q05860 mus musculus
28	132.5	38.7	1362	1 BRD4_HUMAN	O60885 homo sapien
29	132	38.6	415	1 ACRO_PIG	P08001 sus scrofa
30	132	38.6	1790	1 SEPA_EMENI	P78621 emericella
31	131.5	38.5	440	1 G3PT_MOUSE	Q64467 mus musculus
32	131.5	38.5	979	1 RFX1_HUMAN	P22670 homo sapien
33	130.5	38.2	2715	1 TRX2_HUMAN	Q9unn6 homo sapien

RESULT 1

ID	PERT_BORPA	STANDARD;	PRT;	922 AA.
AC	P24328;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.70) (P.95).			
GN	PRN.			
OS	Bordetella pertussis.			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=519;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN2591;			
RX	MEDLINE=91251771; PubMed=2041476;			
RA	Li L.J.; Dougan G.; Novotny P.; Charles I.G.;			
RT	"P.70 pertactin, an outer-membrane protein from Bordetella			
RT	parapertussis; cloning, nucleotide sequence and surface expression in			
RT	Escherichia coli."			
RL	Mol. Microbiol. 5:409-417(1991).			
CC	-!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS			
CC	MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN			
CC	BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-!- SUBUNIT: MONOMER.			
CC	-!- SUBCELLULAR LOCATION: Outer membrane.			
CC	-!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)			
CC	CONCENTRATIONS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X54547; CAA38419.1;			
DR	EMBL; A26124; CAA01786.1;			
DR	EMBL; A19182; CAA01454.1;			
DR	PIR; S15204; S15204.			
DR	PIR; S14659; S14659.			
DR	InterPro; IPR004899; Pertactin_sup.			
DR	InterPro; IPR003992; Pertactin.			
DR	InterPro; IPR003991; Pertactin_vir.			
DR	Pfam; PF03212; Pertactin; 1.			
DR	PRINTS; PR01482; PERTACTIN.			
DR	PRINTS; PR01484; PERTACTINFAMLY.			
KW	Outer membrane; Signal; Virulence; Repeat.			
FT	SIGNAL 1 34 POTENTIAL.			
FT	CHAIN 35 922			
FT	CHAIN 35 647			
FT	PROPEP 648 922			
FT	SITE 260 262			
FT	CELL ATTACHMENT SITE (INVOLVED IN			

P25050 herpesvirus
P04726 triticum ae
P41479 autographa
Q63627 rattus norv
Q95107 bos taurus
Q924h2 mus musculu
Q98av1 arabidopsis
O00401 homo sapien
O10341 orgyia pseu
P04725 triticum ae
P51532 homo sapien
P58840 homo sapien

ALIGNMENTS

DR	Pfam:	pf00657; Lipase_GDSL; 1.
DR	PROSITE;	PS01098; LIPASE_GDSL_SER; 1.
FT	NON_TER	1
FT	ACT_SITE	132 132 BY SIMILARITY.
FT	ACT_SITE	428 428 POTENTIAL.
SQ	SEQUENCE	449 AA; 48779 MW; 9EFB6A3AB28EEA15 CRC64;
Query Match 47.2%; Score 161.5; DB 1; Length 449;		
Best Local Similarity 45.2%; Pred. No. 0.0011;		
Matches 28; Conservative 4; Mismatches 17; Indels 13; Gaps		
QY	3 KAPPAKPAQPQGPGPGS-----POGPGPGPPGPQQPQQPQQPQQPQQPQA	49
DB	SS KPPAPGSPKPGSPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSP	114
QY	50 PP 51	
DB	115 PP 116	
RESULT 5		
ID	YPRO_OWFEU STANDARD; PRT; 141 AA.	
AC	P21260; P21261;	
DT	01-MAY-1991 (Rel. 18, Created)	
DT	01-MAY-1991 (Rel. 18, Last sequence update)	
DT	01-NOV-1995 (Rel. 32, Last annotation update)	
DE	Hypothetical proline-rich protein (Fragment).	
OS	Owenia fusiformis.	
OC	Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canaliculipata;	
OC	Sabellidae; Oweniidae; Owenia.	
ON	NCBI_TaxID=6347;	
RX	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=90147742; PubMed=2105723;	
RA	Bakalaria N., Collet J., Planells R., Thouveny Y., Fontes M.;	
RT	"Presence in invertebrate genomes of sequences characterized by the	
RL	repetition of the triplet CCGurine."	
RL	Biochem. Biophys. Res. Commun. 166:66-73(1990).	
PIR:	A34043; A34043.	
DIR:	B34043; B34043.	
KW	Hypothetical protein; DNA-binding.	
FT	DOMAIN 1 1	
FT	NON_TER 9 58 POLY-PRO.	
FT	DNA_BIND 98 116 H-T-H MOTIF (POTENTIAL).	
FT	NON_TER 141 141	
SQ	SEQUENCE 141 AA; 15745 MW; B294E884D152BDB9 CRC64;	
Query Match 46.2%; Score 158; DB 1; Length 141;		
Best Local Similarity 54.0%; Pred. No. 0.00079;		
Matches 27; Conservative 0; Mismatches 23; Indels 0; Gaps		
QY	5 PPAPKPAQPQGPGPGPGPGPPGPQQPQQPQQPQQPQQPQA	54
DB	10 PPP	59
RESULT 6		
ID	SSGP_VOLCA STANDARD; PRT; 485 AA.	
AC	P21957;	
DT	01-AUG-1991 (Rel. 19, Created)	
DT	01-AUG-1991 (Rel. 19, Last sequence update)	
DT	01-OCT-1996 (Rel. 34, Last annotation update)	
DE	Sulfated surface glycoprotein 185 (SSG 185).	
OS	Volvox carteri.	
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;	
OC	Volvocaceae; Volvox.	
ON	NCBI_TaxID=3067;	
RX	[1]	
RP	SEQUENCE FROM N.A.; AND PARTIAL SEQUENCE.	
RC	STRAIN=f. Nagariensis / HK10;	
RX	MEDLINE=90094551; PubMed=2689458;	

[illegible]

[illegible]


```
FT NON_TER 1 1
SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;

Query Match
Best Local Similarity 42.4%; Score 145; DB 1; Length 296;
Matches 29; Conservative 2; Mismatches 22; Indels 4; Gaps 2;

QY 1 GAKAPPAPKAPQGPQ--GPQGPQGPQGPQGPQ--PPQPPQPPAPAPQPPAG 53
Db 44 GPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPP 100

RESULT 14
ACRO_RABIT STANDARD; PRT; 431 AA.
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
GN ACR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE-Testis;
RX MEDLINE=94368861; PubMed=8086468;
RA Richardson R.T., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
RL Blochim. Biophys. Acta 1219:215-218(1994).
CC -1- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOEA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
CC IS SYNTHESIZED IN A ZMOGEN FORM, PROACROSIN AND STORED IN THE
CC ACROSOME.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -1- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL; U05204; AAA61630.1;
DR HSPP; P00763; IDPO.
DR MEMOPS; S01.223;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16
FT CHAIN 17 431 ACROSIN.
FT CHAIN 17 39 ACROSIN LIGHT CHAIN (BY SIMILARITY).
FT CHAIN 40 ? ACROSIN HEAVY CHAIN (BY SIMILARITY).
FT PROPEP ? 431 PRO-RICH.
FT DOMAIN 40 288 SERINE PROTEASE.
FT DISULFID 22 152 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 160 INTERCHAIN (BY SIMILARITY).
FT DISULFID 71 87 BY SIMILARITY.
FT DISULFID 175 244 BY SIMILARITY.
```

```
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 234 264 BY SIMILARITY.
FT CARBOHYD 19 19 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC... ) (POTENTIAL).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 431 AA; 46422 MW; 1C015A4E0BC0C668 CRC64;

Query Match
Best Local Similarity 42.1%; Score 144; DB 1; Length 431;
Matches 27; Conservative 0; Mismatches 24; Indels 4; Gaps 1;

QY 2 AKAPPAPKAPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGP 52
Db 325 ASGPPPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPP 379

RESULT 15
EXLP_TOBAC STANDARD; PRT; 426 AA.
AC Q03211;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pistil-specific extensin-like protein precursor (PELP).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-cv. Petit Havana; TISSUE-Pistil;
RX MEDLINE=93005740; PubMed=1392607;
RA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;
RT "Developmental expression of tobacco pistil-specific genes encoding
RL novel extensin-like proteins."
CC Plant Cell 4:1041-1051(1992).
CC -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER
CC PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING
CC FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE
CC AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER
CC POLLINATION.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL; Z14019; CAA78397.1;
DR PIR; J01696; J01696.
DR InterPro; IPR000419; Pollen_Ole_e_1.
DR Pfam; PF01190; Pollen_Ole_e_1; 1.
KW Structural protein; Signal; Repeat; Glycoprotein.
FT SIGNAL 1 23
FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.
FT DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).
FT REPEAT 69 73 1.
FT REPEAT 76 80 2.
FT REPEAT 83 87 3.
FT REPEAT 178 182 4.
FT CARBOHYD 310 310
SQ SEQUENCE 426 AA; 44278 MW; 51A495CC94017812 CRC64;

Query Match
Best Local Similarity 41.8%; Score 143; DB 1; Length 426;
Matches 26; Conservative 4; Mismatches 20; Indels 2; Gaps 1;
```

```

QY 2 AKAPAPAPAPQPGPQPGPQPGPQPGPQPGPQPGPQPGPAPAP--Qpp 51
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 AKQPPPPAKPSPPPPPPPVKAPSPSAKQPPPPPPPVKAPSPSAQPP 216

```

Search completed: May 7, 2003, 16:48:08
Job time : 6.72913 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	331.5	96.9	115	2	Q9K5H3	Q9K5H3 bordetella
2	298.5	87.3	111	2	Q9K5G7	Q9K5G7 bordetella
3	294.5	86.1	111	2	Q9K5G9	Q9K5G9 bordetella
4	294.5	86.1	122	2	Q9K7X9	Q9K7X9 bordetella
5	294.5	86.1	252	2	Q9ALQ2	Q9ALQ2 bordetella
6	274	80.1	109	2	Q9K5H1	Q9K5H1 bordetella
7	274	80.1	215	2	Q9ALQ4	Q9ALQ4 bordetella
8	267.5	78.2	107	2	Q9K5H5	Q9K5H5 bordetella
9	267.5	78.2	200	2	Q9ALP9	Q9ALP9 bordetella
10	267.5	78.2	911	2	Q9L4E2	Q9L4E2 bordetella
11	249.5	73.0	107	2	Q9K5G1	Q9K5G1 bordetella
12	249.5	73.0	231	2	Q9ALP3	Q9ALP3 bordetella
13	247	72.2	104	2	Q9K5H9	Q9K5H9 bordetella
14	247	72.2	105	2	Q9K5H7	Q9K5H7 bordetella
15	247	72.2	115	2	Q9KJY1	Q9KJY1 bordetella
16	247	72.2	115	2	Q9KJY0	Q9KJY0 bordetella

RESULT 11	
99K5G1	
ID	PRELIMINARY; PRT; 107 AA.
AC	99K5G1;
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Pertactin (P.68) (Fragment).
OE	PRN
SN	Bordetella bronchiseptica (Alcaligenes bronchisepticus).
NS	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC	

```

Query Match      73.0%; Score 249.5; DB 2; Length 231;
Best Local Similarity 78.0%; Pred. No. 2e-13;
Matches 46; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

          QY      1 GAKAPPAPKAPQGGPQGPGQPQQPPQPPQPPQREAPAPQPPAGRELSAA 59
            ||| ||||| ||||| ||||| ||||| : ||||| ||||| |||||
          Ddb     164 GAKVPPAPKAPQGGPQP-----PQPPQPPQPPQPPQREAPAPQPPAGRELSAA 213

RESULT 13
          Q9K5H9 PRELIMINARY; PRT; 104 AA.
          ID Q9K5H9
          AC Q9K5H9;
```

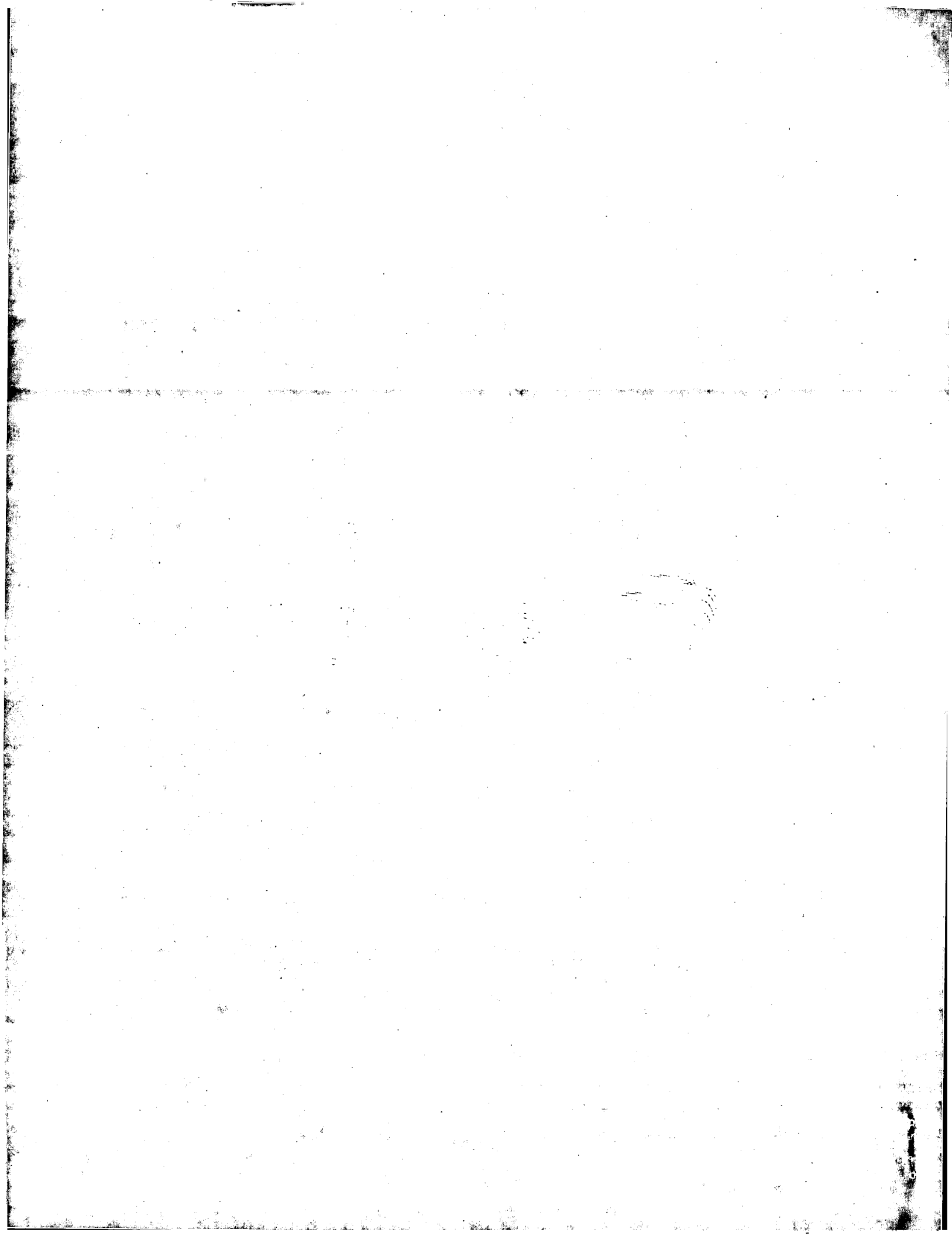
```

AC Q9KYJ1;
DT 01-OCT-2000 ($EMBLrel. 15, Created)
DT 01-OCT-2000 ($FMBLrel. 15, Last sequence update)
DE 01-JUN-2002 ($FMBLrel. 21, Last annotation update)
DE Pertactin (Fragment).
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RA Keil D.J.; Fenwick B.;
RT "variability in the pertactin genes of Bordetella bronchiseptica isolates includes regions coding for the GGXP and POP families of amino acid repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156771; AAF62395.1; -
DR InterPro; IPR004899; Pertact_sup.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PRO1217; PRICHEXTENSN.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 11693 MW; FA3A76596F7097EB CRC64;

Query Match          72.2%; Score 247; DB.2; Length 115;
Best Local Similarity 80.0%; Pred. No. 1.8e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 12; Gaps 90

QY      1 GAKAPPAPKPAQPQGPGPQGPGGPPGPQQPPQPQR-PEAPAPPPAGRELSAA 95
       |||||
Db     42 GAKAPPAPKPAQPQGPGG-----PQPPQPQPQPQRPEAPAPPPAGRELSAA 90

Search completed: May 7, 2003, 16:53:13
Job time : 30.1533 secs
```



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 27.0187 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-17
Perfect score: 298
Sequence: 1 GAKAPPAPKAPQPGQPGP.....PORPEAPAPQPPAGRELSAA 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	287.5	96.5	53	23	AAE16196 B. bronchiseptica
2	282	94.6	56	23	AAE16197 B. bronchiseptica
3	274	91.9	60	23	AAE16195 B. bronchiseptica
4	271	90.9	52	23	AAE16194 B. bronchiseptica
5	271	90.9	911	12	AAAR14320 Pertactin antigen
6	271	90.9	911	13	AAAR26503 prn proteins. Bor
7	271	90.9	911	23	AAE16183 Bordetella bronchi
8	267	89.6	58	23	AAE16198 B. bronchiseptica
9	267	89.6	922	13	AAAR25578 Bordetella parapap
10	267	89.6	922	23	AAE16185 Bordetella parapap

11	258	86.6	922	12	AAAR14321	Pertactin antigen
12	250.5	84.1	49	23	AAE16193	B. bronchiseptica
13	244	81.9	48	23	AAE16199	B. bronchiseptica
14	236	79.2	52	23	AAE16200	B. bronchiseptica
15	235	78.9	54	23	AAE16201	B. bronchiseptica
16	207	69.5	42	23	AAE16202	B. bronchiseptica
17	207	69.5	910	23	AAE16184	Bordetella pertuss
18	207	69.5	910	23	AAE17146	Bordetella pertuss
19	186.5	62.6	39	23	AAE16203	B. bronchiseptica
20	153	51.3	31	12	AAE11737	Peptide 683 deriv
21	149	50.0	31	12	AAE11739	Peptide B805 deriv
22	136.5	45.8	446	22	ABG70063	Drosophila melanog
23	134.5	45.1	406	22	ABG27250	Novel human diago
24	134	45.0	80	22	AAO04412	Human polypeptide
25	134	45.0	99	22	AAO02076	Human polypeptide
26	133.5	44.8	900	21	ABAR2321	Human ORFX ORF2085
27	133.5	44.8	1217	22	ABG09876	Novel human diago
28	133.5	44.8	1239	22	ABG09877	Novel human diago
29	133	44.6	85	22	AAO04573	Human polypeptide
30	132	44.3	350	23	ABG93838	Human polypeptide
31	131	44.0	439	13	AAAR28150	Herbicidally activ
32	130	43.6	206	18	AAAR14574	Sugar beet chitina
33	129	43.3	88	22	AAO02124	Streptococcus pneu
34	129	43.3	183	18	AAAR14570	Human polypeptide
35	128.5	43.1	76	23	ABG78337	Streptococcus pneu
36	128	43.0	129	18	AAAR14573	Ser-Pro-Pro-Pro ge
37	128	43.0	190	18	AAAR14569	Streptococcus pneu
38	128	43.0	358	21	AAAG22949	Streptococcus pneu
39	127.5	42.8	85	22	AAO02036	Arabidopsis thalia
40	127	42.6	325	22	ABG21919	Human polypeptide
41	126.5	42.4	470	21	ABG49660	Novel human diago
42	126	42.3	598	22	ABG14000	Arabidopsis thalia
43	125	41.9	24	17	AAAR87649	Novel human diago
44	124.5	41.8	329	23	AAE14305	Synthetic Bordetel
45	123.5	41.4	356	22	ABG04360	Sugarcane proline
						Novel human diago

ALIGNMENTS

RESULT 1
AAE16196

ID AAE16196 standard; peptide; 53 AA.

AC AAE16196;

DT 26-MAR-2002 (first entry)

XX B. bronchiseptica strain II-4 pertactin outer membrane protein region II.

XX Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX therapy; antibiotic; antibacterial; region II.

OS Bordetella bronchiseptica.

PN WO200190143-A2.

PD 29-NOV-2001.

PF 23-MAY-2001; 2001WO-EP06457.

PR 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of
XX pertactin in Bordetella species, useful in immunogenic compositions for
XX treating infections caused by Bordetella and in diagnostic methods

XX	28-MAR-1991;	91WO-G800487.
XX	02-APR-1990;	90GB-0007416.
XX	(WELL)	WELLCOME FOUNDATION LTD.
XX	Clare JJ,	Romanos MA;
XX	WPI;	1991-325214/44.
XX	N-PSDB;	AAQ14320.
XX	Pichia microorganism transformants - for production of	
PT	Bordetella pertactin antigens for whooping cough vaccines	
XX	Disclosure; Fig 1C; 38pp; English.	
XX	Pichia microorganisms are transformed for the expression of	
CC	pertactin antigens. DNA sequence used are represented in SEQ	
CC	encoding the B. bronchiseptica P.68 and B. pertussis P.70/70	
CC	respectively or the B. pertussis P.69 encoding sequence descri	
CC	by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80: 35	
CC	C (1989).	
CC	The 46 Xs represent amino acids missing in the specification.	
XX	Sequence	922 AA;
XX	Query Match	86.6%; Score 258; DB 12; Length 922;
XX	Best Local Similarity	86.4%; Pred. No. 2.3e-12;
XX	Matches 51;	Conservative 0; Mismatches 0; Indels 8
OY	1 GAKAPPAPKAPGPGPGPGQPQ- - - - - PQPPQPQR-PEAPAPQPPAGRELL	
Db	564 GAKAPPAPKAPGPGPGQPQ- - - - - PQPPPPXXPPQPPQPPQPEAPAPQPPAGRELL	
XX	RESULT 12	
XX	AAEL16193	
ID	AAEL16193 standard; peptide; 49 AA.	
XX	AAEL16193;	
XX	26-MAR-2002 (first entry)	
XX	B. bronchiseptica strain II-1 pertactin outer membrane protein	
DE	Pertactin; PRN; outer membrane protein; vaccine; Bordetella i	
XX	therapy; antibiotic; antibacterial; region II.	
XX	Bordetella bronchiseptica.	
OS	WO200190143-A2.	
XX	29-NOV-2001.	
XX	23-MAY-2001; 2001WO-EP06457.	
XX	25-MAY-2000; 2000US-206969P.	
XX	(INSP) INST PASTEUR.	
XX	Guiso-maclouf N, Boursaux-eude C;	
XX	WPI; 2002-097639/13.	
XX	Polypeptides containing polymorphisms of the repeated regions	
PT	pertactin in Bordetella species, useful in immunogenic composi	
PT	treating infections caused by Bordetella and in diagnostic met	
XX	Claim 26; Fig 1c; 47pp; English.	
XX	The present invention relates to Bordetella bronchiseptica per	
CC	(outer membrane protein) or their fragments. Pertactin (PRN) i	

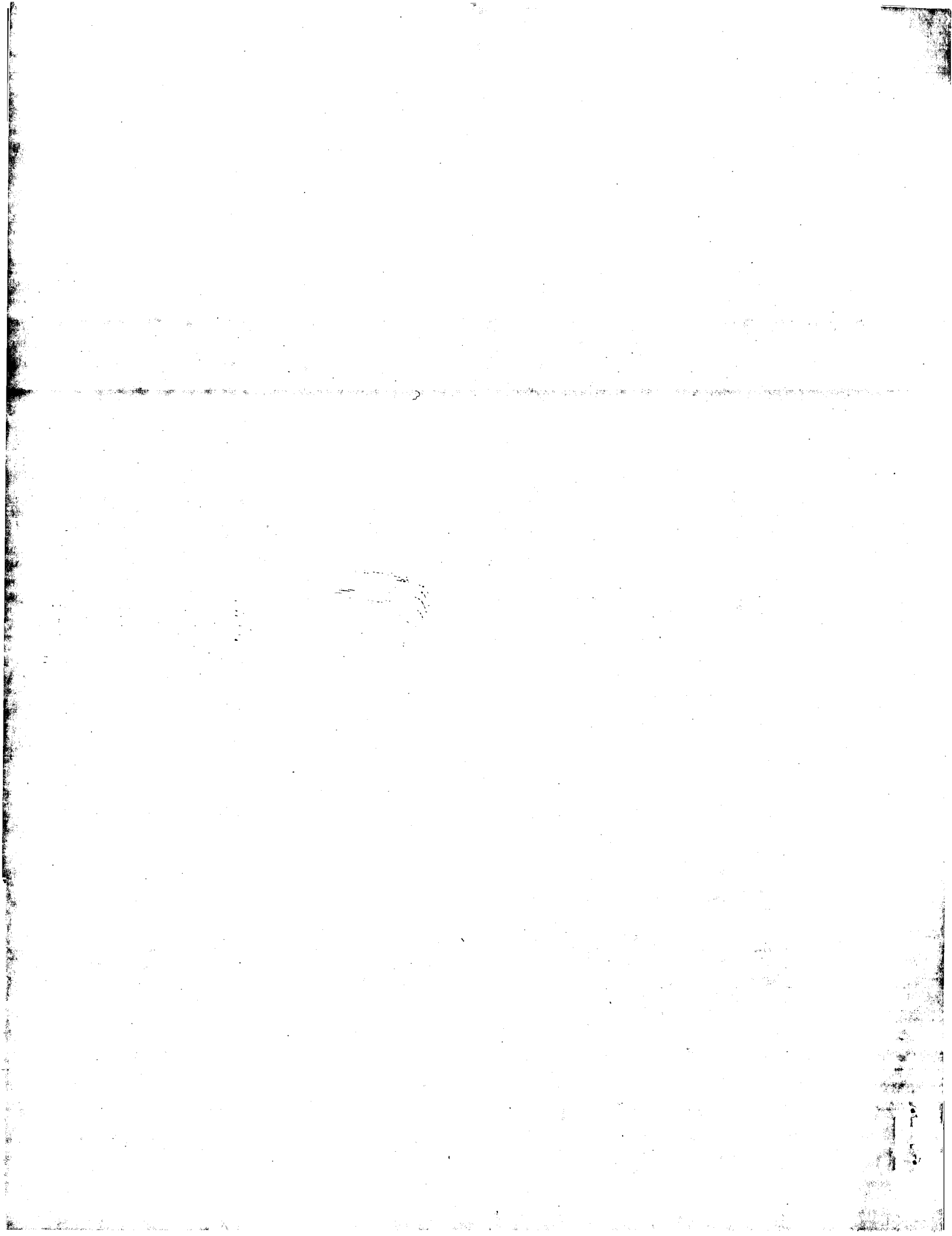
AAE16201
ID AAE16201 standard; peptide: 54 AA.

AE16201
ID AE16201 standard; peptide: 54 AA.

QY		1 GAKAPPKPAQPQGPGPQGPPGP-PEAPAPPAGRELSAA	52
Db		559 GAKAPPKPAQPQGPGPOP-POPPOPPORPEAPAPPAGRELSAA	610

RESULT 9
US-08-453-265-16

Query Match 40.6%; Score 121; DB 6; Length 331;
Best Local Similarity 47.7%; Pred. No. 0.00053;
Matches 21; Conservative 2; Mismatches 21; Indels



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	271	90.9	911	2	A47675	68K outer membrane	
2	267	89.6	922	2	S15204	pactactin - Bordet	
3	207	69.5	910	2	A32560	outer membrane pro	
4	145.5	48.8	430	2	JC2301	hypothetical 47.8K	
5	145	48.7	209	2	A48232	cysteine-rich exte	
6	140.5	47.1	449	2	S16748	proline-rich prote	
7	138.5	46.5	446	2	T07907	hydroxyproline-ric	
8	138	46.3	196	2	B48232	cysteine-rich exte	
9	138	46.3	270	2	H83619	hypothetical prote	
10	137	46.0	383	2	AE2295	hypothetical prote	
11	137	46.0	599	2	T10798	phosphorin-S - Vo	
12	136	45.6	485	2	A33647	sulfated surface g	
13	135	45.3	487	2	S42442	nuclear protein BB	
14	134	45.0	141	2	A34043	hypothetical proli	
15	132.5	44.5	3164	1	WMBEH6	UL36 protein - hum	
16	132	44.3	534	2	S21961	proline-rich prote	
17	131	44.0	439	2	S51939	chitinase (EC 3.2.	
18	130	43.6	288	2	T17737	proline-rich prote	
19	130	43.6	1137	2	A86335	T20H2.9 protein -	
20	130	43.6	1494	2	T14355	proline-tyrosine-p	
21	129.5	43.5	165	2	T24470	hypothetical prote	
22	129.5	43.5	412	2	T17531	proline-rich prote	
23	129.5	43.5	875	2	T10340	hypothetical prote	
24	129	43.3	1952	2	T48814	hypothetical prote	
25	127.5	42.8	1496	2	T17908	proline/lysine-ric	
26	126.5	42.4	1684	2	T02632	hypothetical prote	
27	126	42.3	544	2	T17547	proline-rich prote	
28	125.5	42.1	339	2	T17636	proline-rich prote	
29	125	41.9	225	2	T17815	proline-rich prote	

[illegible]

RESULT 8

cysteine-rich extensin-like protein 2 precursor - common tobacco
 C:Species: *Nicotiana tabacum* (common tobacco)
 C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 01-Dec-2000
 C:Accession: B48232; PQ0474; S24616
 R:Wu, H.; Zou, H.; May, B.; Gu, Q.; Cheung, A.Y.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6829-6833, 1993
 A:Title: A tobacco gene family for flower cell wall proteins with a proline-rich domain
 A:Reference number: A48232; MUID:93342083; PMID:8341705
 A:Accession: B48232
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-196 <WUA>
 A:Cross-references: GB:L13440; NID:g310924; PIDN:AAA34060.1; PID:g310925
 E:de S Goldman, M.H.; Pezzottoli, M.; Seurinck, J.; Mariani, C.
 Plant Cell 4, 1041-1051, 1992
 A:Title: Developmental expression of tobacco p1st1-specific genes encoding novel extensins
 A:Reference number: PQ0474; MUID:93005740; PMID:1392607

A:Accession: PQ0474
A:Accession: PQ0474
A:Molecule type: mRNA
A:Residues: 'NAG', 1-105 <GOL>
A:Cross-references: EMBL:Z14014
A:Experimental source: stigma; strain Petit Havana SRI
C:Genetics:
C:Gene: CENP-2
C:Superfamily: glutelin
C:Keywords: cell wall; extracellular matrix; fertilization
F:10/Domain: signal sequence #status predicted <SIG>
F:20-136/Product: cysteine-rich extensin-like protein 2 #stat

Query Match	46.3%	Score 138;	DB 2;	Length 196;
Best Local Similarity	55.0%	Pred. No. 0.0035;		
Matches 22; Conservative		3; Mismatches 15;	Indels 0;	Gaps 0;

Qy	6	PAPKAPQPQGPGPGPQPGPQPPQPPQPPQPPQPPQPPA	45
		:	
Db	71	PRRRCPSPPPPRRRRCPSPPPPQPPRPRSPPPPPPPPS	110

RESULT 9

H83619
hypothetical protein PA0197 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83619
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: AB2950; MID:20437337; PMID:10984043
A:Accession: H83619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <STO>
A:Cross-references: GB:AE004458; GB:AE004091; NID:G9946031; PIDN:AAG03586.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
C:Gene: PA0197
C:Superfamily: tonB protein

Query Match 46.3%; Score 138; DB 2; Length 270;
Best Local Similarity 52.0%; Pred. No. 0.0045;
Matches 26; Conservative 3; Mismatches 15; Indels

Qy 2 AKAPPAPKPAQPGPQPGPQP-----PQPPQPPQPPQPPA 45

Db . 123 AVKPPPKPVKPKPKPKQPRPKAPKAVEFAFPQPQAAPPAPA 172

RESULT 10

hypothetical protein all3916 [Imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc 'sp.'
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-
C:Accession: AE2295
R:kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Wata-
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yae
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2295
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-383 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA875615.1; PID:gi7133050; GSPDB
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3916

```

Query Match          46.08; Score 137; DB 2; Length 383;
Best Local Similarity 55.08; Pred. No. 0.0069;
Matches 22; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 5 PPAPKPAFQCGPGQPGFGQPPQPPQPPQPPQPPQPPQPP 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 PPPDDPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11
T10798

i10750
 perophorin-S - Völvox carteri
 C;Species: Völvox carteri
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C;Accession: T10798
 R;Godl, K.; Hallmann, A.; Wenzl, S.; Sumper, M.
 EMBO J. 16, 25-34, 1997
 A;Title: Differential targeting of closely related ECM-glycoproteins: The perophorin
 A;Reference number: Z17154; MUID: 97162277; PMID: 9009264
 A;Accession: T10798
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-599 <GD>
 A;Cross-references: EMBL:Y07752; NID:g1655698; PIDN:CAA69032.1; PID:g1655699
 A;Experimental source: strain HK 10; sub_species Nagariensis
 A;Note: in contrast to the other perophorins, perophorin-S is targeted to the cell-
 C;Keywords: extracellular matrix; glycoprotein; pheromone

```
Query Match      46.0%   Score 137; DB 2; Length 599;
Best Local Similarity 55.0%; Pred. No. 0.0097;
Matches 22; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
```

RESULT 12

RESULT 12 :
A33647
sulfated surface glycoprotein 185 - Volvox carteri
C;Species: Volvox carteri
C;Date: 11-Apr-1990 #sequence_revision 11-Apr-1990 #text_change 21-Jul-2000
C;Accession: A33647
R;Ertl, H.; Mengele, R.; Wenzl, S.; Engel, J.; Sumper, M.
J. Cell Biol. 109 : 3493-3501, 1989
A;Title: The extracellular matrix of Volvox carteri: molecular structure of the cellu
A;Reference number: A33647; PMID:9009451; PMID:2689458
A;Accession: A33647

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			ID	Description
	Score	Match	Length		
1	271	90.9	911	1	PERT_BORBR
2	267	89.6	922	1	PERT_BORPA
3	207	69.5	910	1	PERT_BORPE
4	140.5	47.1	449	1	APG_BRANA
5	136	45.6	485	1	SSGP_VOLCA
6	135	45.3	487	1	ENR2_EBV
7	134	45.0	141	1	YPRO_OWEFU
8	132.5	44.5	3164	1	TGCU_HSV11
9	130.5	43.8	555	1	GPI_CHLRE
10	130	43.6	534	1	APC_ARATH
11	129.5	43.5	875	1	Y066_NFVOP
12	125	41.9	2004	1	MOZ_HUMAN
13	124.5	41.8	339	1	CSP_PLABE
14	124.5	41.8	347	1	CSP_PLABA
15	124.5	41.8	431	1	ACRO_RABIT
16	124	41.6	102	1	COLL_HSVCT
17	123.5	41.4	426	1	EXLP_TOBAC
18	122	40.9	415	1	ACRO_FIG
19	121.5	40.8	261	1	PRP2_MOUSE
20	121.5	40.8	296	1	PRP3_MOUSE
21	121.5	40.8	565	1	MOT8_MOUSE
22	120	40.3	440	1	G3PT_MOUSE
23	120	40.3	1790	1	SEPA_EMENI
24	118.5	39.8	129	1	PARB_TRYBB
25	118.5	39.8	979	1	RFXL_HUMAN
26	118	39.6	105	1	COLL_HSV7
27	118	39.6	380	1	VASP_HUMAN
28	118	39.6	1059	1	CAPU_DROME
29	116.5	39.1	143	1	PAR1_TRYBB
30	116.5	39.1	145	1	PARC_TRYBB
31	116.5	39.1	443	1	HXA3_MOUSE
32	116	38.9	232	1	ACRL_HUMAN
33	116	38.9	1206	1	FW14_MOUSE

```

ADHESION TO VARIOUS EUKARYOTIC CELL LINES).
4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
1.
2.
3.
4. (APPROXIMATE).
9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
SEQUENCE 922 AA; 95178 MW; 3DF/BF58D47112478 CRC64;

Query Match      89.6%; Score 267; DB 1; Length 922;
Best Local Similarity 87.9%; Pred. No. 2.8e-10;
Matches 51; Conservative 0; Mismatches 1; Indels 6; Gaps 2;

QOY 1 GAKAPPAPKPAPQGPGQP-----GQQPPPPPPOR-PEARAPQPPAGRELSAA 52
|||||
DBD 564 GAKAPPAPKPAPQGPGQPFPQPQPQPQPQPQPQPQPQPPEANPAPQPPAGRELSAA 621
|||||

RESULT 3
PERT_BORPE STANDARD; PRT; 910 AA.
ID PERT_BORPE AC PI4283;
DC 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DD 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
DN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RZ SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-CN2992;
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RN [2]
RZ REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica.";
RL J. Gen. Microbiol. 138:1697-1705(1992).
RN [3]
RZ X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RL Nature 381:90-92(1996).
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW Mg(2+)
CC CONCENTRATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04560; AAA22980.1; ALT SEQ.

```

```

Query Match          45.6%; Score 136; DB 1; Length 485;
Best Local Similarity 55.0%; Pred. No. 0.016;
Matches 22; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY      5 PPAPKPAFQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQPP 44
      || | | | | | | | | | | | | | | | | | | |
Db      261 PPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPSPSP 300

RESULT 6
EBN2_EBV
ID      EBN2_EBV          STANDARD;          PRT;      487 AA.
AC      P12978;

```


Search completed: May 7, 2003, 16:48:09
Job time : 6.0494 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 29.0971 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754b-18
Perfect score: 322
Sequence: 1 GAKAPPAPKAPQPGQPGP.....ORQPEAPAPQPPAGRELSAA 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	322	100.0	56	23	AAE16197
2	310	96.3	60	23	AAE16195
3	303	94.1	58	23	AAE16198
4	303	94.1	922	13	AAE25578
5	303	94.1	922	13	AAE16185
6	294	91.3	922	12	AAE14321
7	291.5	90.5	53	23	AAE16196
8	285	88.5	52	23	AAE16194
9	285	88.5	911	12	AAE14320
10	285	88.5	911	13	AAE26503

11	285	88.5	911	23	AAE16183	Bordetella bronch
12	264.5	82.1	49	23	AAE16193	B. bronchiseptica
13	258	80.1	48	23	AAE16199	B. bronchiseptica
14	257	79.8	52	23	AAE16200	B. bronchiseptica
15	256	79.5	54	23	AAE16201	B. bronchiseptica
16	210	65.2	42	23	AAE16202	B. bronchiseptica
17	210	65.2	910	23	AAE16184	Bordetella pertuss
18	210	65.2	910	23	AAE17146	Bordetella pertuss
19	189.5	58.9	39	23	AAE16203	B. bronchiseptica
20	155.5	48.3	900	21	AAE2321	Human OREF ORF2085
21	155.5	48.3	1217	22	ABG09876	Novel human diagno
22	155.5	48.3	1239	22	ABG09877	Novel human diagno
23	155	48.1	31	12	AAE11737	Peptide 683 derive
24	152	47.2	31	12	AAE11739	Peptide BBO5 deriv
25	146	45.3	80	22	AAO04412	Human polypeptide
26	146	45.3	99	22	AAO02076	Human polypeptide
27	145	45.0	85	22	AAO04573	Human polypeptide
28	145	45.0	446	22	ABE70063	Drosophila melanog
29	142	44.1	88	22	AAO02124	Human polypeptide
30	141.5	43.9	76	23	ABE78537	Ser-Pro-Pro-Pro ge
31	141.5	43.9	406	22	ABG27250	Novel human diagno
32	140	43.5	439	13	AAE28150	Sugar beet chitina
33	140	43.5	470	21	AAE49660	Arabidopsis thalia
34	138	42.9	85	22	AAO02036	Human polypeptide
35	138	42.9	325	22	ABG21919	Novel human diagno
36	138	42.9	350	23	ABE93838	Herbicidally activ
37	134	41.6	334	9	AAE82971	Bioadhesive precu
38	134	41.6	334	9	AAE83194	Sequence of a bio
39	134	41.6	358	21	AAE22949	Arabidopsis thalia
40	133.5	41.5	993	22	ABE64403	Drosophila melanog
41	133	41.3	538	22	AAE82806	Human low density
42	132.5	41.1	550	22	AAE82807	Rabbit low density
43	132	41.0	206	18	AAE14574	Streptococcus pneu
44	130.5	40.5	104	23	ABE78538	Ser-Pro-Pro-Pro-Pr
45	130.5	40.5	594	22	ABE61362	Drosophila melanog

ALIGNMENTS

RESULT 1
AAE16197

ID AAE16197 standard; peptide; 56 AA.

AC AAE16197;

DT 26-MAR-2002 (first entry)

DE B. bronchiseptica strain II-5 pertactin outer membrane protein region II.
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella Infection;
KW therapy; antibiotic; antibacterial; region II.

OS Bordetella bronchiseptica.

PN WO200190143-A2.

PD 29-NOV-2001.

PF 23-MAY-2001; 2001WO-EP06457.

PR 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

[illegible]

RESULT 14
AAE16200
ID AAE16200 standard; peptide; 52 AA.

XX	
AC	AAE16200;
XX	
DT	26-MAR-2002 (first entry)

XX
DE B. bronchiseptica strain II-8 pertactin outer membrane protein region II.
XX
KW Pertactin: PRN: outer membrane protein: vaccine. Bordetella infection:

XX OS Bordetella bron
XX PN WO200190143-A2

XX	29-NOV-2001.
PD	
XX	
PF	23-MAY-2001: 2001WO-EP06457

XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP \ INST PASTEUR

XX
PI Guiso-maclouf N, Boursaux-eude C;
XX

DR WPT; 2002-09/639/13.

polypeptides containing poly

PT treating infections caused by

PS Claim 26; Fig. 1c; 47pp; Engl

CC The present invention

CC vaccine. Pertactin antibody is used

CC and used to detect
CC purifying correspond

used in affinity chromatography to identify antibodies

CC animal tissue and human or other animal cells, as well as biological
CC to identify antibodies to sorbitella in materials such as human or other
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the

antigens can be used for qualitative

CC	XX	SQ	Sequence	Score	DB	Length
CC	XX	SQ	perfactin outer membrane protein region II.	79.8%	257	52

AAE16201
ID AAE16201 standard: pentide: 54 AA

[illegible]

RESULT 15
AAE16201
ID AAE16201 standard: pentide: 5A AA

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	303	94.1	922	4	US-08-460-269C-6	Sequence 6, Appli
2	285	88.5	911	4	US-08-460-269C-4	Sequence 4, Appli
3	210	65.2	910	4	US-08-460-269C-2	Sequence 2, Appli
4	136	42.2	33	1	US-08-237-716-11	Sequence 11, Appli
5	134	41.6	331	6	5202236-37	Patent No. 5202236
6	134	41.6	334	6	5202236-3	Patent No. 5202236
7	127	39.4	24	4	US-08-750-624-11	Sequence 11, Appli
8	123.5	38.4	3119	1	US-08-246-982A-16	Sequence 16, Appli
9	123.5	38.4	3119	1	US-08-453-265-16	Sequence 16, Appli
10	122.5	38.0	1274	4	US-09-095-443-2	Sequence 2, Appli
11	121.5	37.7	478	3	US-08-155-888-2	Sequence 2, Appli
12	121.5	37.7	513	4	US-09-041-886-28	Sequence 28, Appli
13	121.5	37.7	530	4	US-09-041-886-29	Sequence 29, Appli
14	121.5	37.7	552	4	US-09-041-886-30	Sequence 30, Appli
15	121.5	37.7	589	4	US-09-041-886-31	Sequence 31, Appli
16	121.5	37.7	3144	1	US-08-246-982A-6	Sequence 6, Appli
17	121.5	37.7	3144	1	US-08-453-265-6	Sequence 6, Appli
18	121.5	37.7	3144	2	US-08-457-273B-42	Sequence 42, Appli
19	121.5	37.7	3144	3	US-08-556-419-21	Sequence 21, Appli
20	121.5	37.7	3144	4	US-09-041-886-15	Sequence 15, Appli
21	120.5	37.4	106	4	US-09-314-268-134	Sequence 134, App
22	120.5	37.4	490	4	US-09-109-841-2	Sequence 2, Appli
23	120.5	37.4	641	4	US-08-961-083-160	Sequence 160, App
24	120	37.3	23	4	US-08-460-269C-8	Sequence 8, Appli
25	115.5	35.9	3118	2	US-08-457-273B-8	Sequence 8, Appli
26	115	35.7	330	1	US-08-642-255-32	Sequence 32, Appli
27	115	35.7	408	1	US-07-609-716-65	Sequence 65, Appli

•

;;
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California

```
;;  
;;SEQUENCE CHARACTERISTICS:  
;;LENGTH: 530 amino acids  
;;TYPE: amino acid  
;;TOPOLOGY: linear
```

, GENERAL INFORMATION:
 , APPLICANT: Bredesen, Dale E.
 , APPLICANT: Rabizadeh, Sharroz
 , TITLE OF INVENTION: Proapoptotic Peptides, Dependence
 , TITLE OF INVENTION: Polypeptides and Methods of Use
 , NUMBER OF SEQUENCES: 72
 , CORRESPONDENCE ADDRESS:


```

DR SMART; SW00352; POU; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; Nuclear protein; DNA-binding; Homeobox;
KW Developmental protein.
FT DOMAIN 57 66 POU-IV BOX.
FT DOMAIN 100 108 POLY-HIS.
FT DOMAIN 130 189 GLY-RICH.
FT DOMAIN 207 260 ALA-RICH.
FT DOMAIN 267 341 POU.
FT DNA_BIND 359 418 HOMEBOX.
FT CONFLICT 91 92 TS -> H (IN REF. 3).
FT CONFLICT 99 99 A -> R (IN REF. 1).
FT CONFLICT 133 135 GAG -> ARR (IN REF. 3).
FT CONFLICT 135 137 MISSING (IN REF. 2).
FT CONFLICT 139 142 MISSING (IN REF. 3).
FT CONFLICT 149 150 GP -> AA (IN REF. 3).
FT CONFLICT 155 156 GP -> PR (IN REF. 3).
FT CONFLICT 174 175 GP -> AA (IN REF. 3).
FT CONFLICT 189 189 G -> A (IN REF. 3).
FT CONFLICT 240 240 A -> S (IN REF. 2).
FT CONFLICT 252 252 MISSING (IN REF. 3).
FT CONFLICT 256 256 A -> R (IN REF. 3).
FT CONFLICT 276 278 AER -> GS (IN REF. 3).
FT CONFLICT 343 345 AQR -> PS (IN REF. 3).
SQ SEQUENCE 423 AA; 42939 MW; B3AE4732E1309F34 CRC64;

Query Match 24.6%; Score 67; DB 1; Length 423;
Best Local Similarity 56.0%; Pred. No. 7.4;
Matches 14; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 8 GDAPGAVPGGAVPGGAVPGGFGP 32
DB 153 GGGPGGGGPGGGPGGG--GGGGP 175

RESULT 10
BR3A_MOUSE
ID BR3A_MOUSE STANDARD; PRT; 421 AA.
AC P17208;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUN-2002 (Rel. 34, Last sequence update)
DE Brain-specific homeobox/POU domain protein 3A (BRN-3A) (BRN-3.0).
GN POU4F1 OR BRN3A OR BRN3 OR BRN-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94215319; PubMed=8162704;
RA Thell T., Zechner U., Klett C., Adolph S., Moersey T.;
RT "Chromosomal localization and sequences of the murine Brn-3 family of
developmental control genes.";
RL Cytogenet. Cell Genet. 66:267-271(1994).
[2]
RP SEQUENCE OF 286-401 FROM N.A.
RC STRAIN-T6 / TW1; TISSUE=Testis;
RX MEDLINE-90221898; PubMed=1970171;
RA Goldsborough A., Ashworth A., Willison K.;
RT "Cloning and sequencing of POU-boxes expressed in mouse testis.";
RL Nucleic Acids Res. 18:1634-1634(1990).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN
THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF
NEURONAL LINEAGES. A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
CC THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: BRAIN, PERIPHERAL SENSORY NERVOUS SYSTEM AND
RETINA. IN THE ADULT NERVOUS SYSTEM BRN-3.0 PREDOMINATES IN THE
MEDIAL HABENULA, SUPERFICIAL GRAY OF THE SUPERIOR COLLICULUS, RED

```

```

CC NUCLEUS, MESENCEPHALIC NUCLEUS OF THE TRIGEMINAL GANGLION, NUCLEUS
CC AMBIGUUS, INFERIOR OLIVARY NUCLEUS, AND PERIPHERAL SENSORY
CC GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.
CC CLASS-4 SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S69350; AAB30577.2; -
CC EMBL; X51959; CAA36218.1; -
CC PIR; S09237; S09237.
CC HSSP; P10037; 1AU7.
CC MGD; MGI:102525; Pou4f1.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR000327; POU_domain.
CC Pfam; PF00046; Homeobox; 1.
CC Pfam; PF00157; pou; 1.
CC PRINTS; PR00028; POUDOMAIN.
CC ProDom; PD00010; Homeobox; 1.
CC ProDom; PD000583; POU_domain; 1.
CC SMART; SM00389; HOX; 1.
CC SMART; SM00352; POU; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00035; POU_1; 1.
CC PROSITE; PS00465; POU_2; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; Nuclear protein; DNA-binding; Homeobox;
KW Developmental protein.
FT DOMAIN 57 66 POU-IV BOX.
FT DOMAIN 100 108 POLY-HIS.
FT DOMAIN 130 187 GLY-RICH.
FT DOMAIN 205 258 ALA-RICH.
FT DOMAIN 265 339 POU.
FT DNA_BIND 357 416 HOMEBOX.
SQ SEQUENCE 421 AA; 42781 MW; 34EC99D789BEE939 CRC64;

Query Match 24.1%; Score 65.5; DB 1; Length 421;
Best Local Similarity 58.3%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 9 DAPAGGAVP-GGAVPGGAVPGGFG 31
DB 145 DGGGGGGGGGGGGGGGGGGGG 168

RESULT 11
PTPO_RAT
ID PTPO_RAT STANDARD; PRT; 1711 AA.
AC Q64612;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Osteotesticular protein tyrosine phosphatase precursor (EC 3.1.3.48)
DE (OST-PTP).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Osteosarcoma;
RX MEDLINE-95074080; PubMed=7527035;
RA Mauro L.J., Olmsted E.A., Skrobacz B.M., Mourey R.J., Davis A.R.,
RA Dixon J.E.;
RT "Identification of a hormonally regulated protein tyrosine
phosphatase associated with bone and testicular differentiation.";

```

RL J. Biol. Chem. 269:30659-30667(1994).

CC -1- FUNCTION: MAY FUNCTION IN SIGNALING PATHWAYS DURING BONE

CC REMODELING, AS WELL AS SERVE A BROADER ROLE IN CELL INTERACTIONS

CC ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS. OPTIMAL PH FOR

CC PHOSPHATASE ACTIVITY IS 5.6.

CC -1- ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS.

CC CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H₂O -> protein

CC tyrosine + phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be

CC produced by alternative splicing. A presumed alternate transcript

CC of 4.8-5.0 kilobases, which may lack PTP domains, is present in

CC proliferating osteoblasts, but not detectable at other stages.

CC -1- TISSUE SPECIFICITY: BONE AND TESTIS. IN THE LAYER, RESTRICTED TO

CC THE BASAL PORTION OF THE SEMINIFEROUS TUBULE.

CC -1- DEVELOPMENTAL STAGE: UP-REGULATED IN DIFFERENTIATING CULTURES OF

CC PRIMARY OSTEOBLASTS AND DOWN-REGULATED IN LATE STAGE MINERALIZING

CC CULTURES. IN TESTIS, EXPRESSION IS HIGHEST BETWEEN STAGES I AND

CC VII WHEN MATURING SPERMATIDS REMAIN BURIED WITHIN THE SERTOLI

CC EPITHELIUM.

CC -1- INDUCTION: BY PARATHYROID HORMONE AND CYCLIC AMP ANALOGS.

CC -1- PTM: THE CYTOPLASMIC DOMAIN CONTAINS POTENTIAL PHOSPHORYLATION

CC SITES.

CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC -1- SIMILARITY: CONTAINS 10 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.

CC -----

CC EMBL: L36884; AAA63911.1; -

CC HSSP: P18052; 1YFO.

CC DR InterPro: IPR000340; DS_Phosphatase.

CC DR InterPro: IPR003961; FN_III.

CC DR InterPro: IPR003962; FNIII-repeat.

CC DR InterPro: IPR000387; TYR_phosphatase.

CC DR InterPro: IPR000242; TYR_PP.

CC DR Pfam: PF00041; fn3; 7.

CC DR Pfam: PF00102; Y_phosphatase; 1.

CC DR PRINTS: PR00014; FNTYPEIII.

CC DR PRINTS: PR00700; PRTYPHPHTASE.

CC DR SMART: SM00060; FN3; 6.

CC DR SMART: SM00194; PTPC; 1.

CC DR SMART: SM00012; PTPC_DSPC; 1.

CC DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.

CC DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.

CC DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.

CC DR Hydrolase; Transmembrane; Repeat; Signal; Glycoprotein.

CC SIGNAL 1 17 POTENTIAL.

CC CHAIN 18 1711 OSTEOCYTICULAR PROTEIN TYROSINE

CC PHOSPHATASE.

CC FT DOMAIN 18 1074 EXTRACELLULAR (POTENTIAL).

CC FT TRANSMEM 1075 1095 POTENTIAL.

CC FT DOMAIN 1096 1711 CYTOPLASMIC (POTENTIAL).

CC FT DOMAIN 32 124 FIBRONECTIN TYPE-III 1.

CC FT DOMAIN 125 215 FIBRONECTIN TYPE-III 2.

CC FT DOMAIN 216 303 FIBRONECTIN TYPE-III 3.

CC FT DOMAIN 304 392 FIBRONECTIN TYPE-III 4.

CC FT DOMAIN 393 470 FIBRONECTIN TYPE-III 5.

CC FT DOMAIN 471 562 FIBRONECTIN TYPE-III 6.

CC FT DOMAIN 563 652 FIBRONECTIN TYPE-III 7.

CC FT DOMAIN 653 741 FIBRONECTIN TYPE-III 8.

CC FT DOMAIN 742 830 FIBRONECTIN TYPE-III 9.

CC FT DOMAIN 831 921 FIBRONECTIN TYPE-III 10.

CC FT DOMAIN 1150 1418 PROTEIN-TYROSINE PHOSPHATASE 1.

CC FT DOMAIN 1419 1711 PROTEIN-TYROSINE PHOSPHATASE 2.

CC FT ACT_SITE 1350 1350 BY SIMILARITY.

CC FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 982 982 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1711 AA; 187292 MW; 0F04D2D1A47A18A0 CRC64;

Query Match 24.1%; Score 65.5; DB 1; Length 1711;

Best Local Similarity 38.5%; Pred. No. 37;

Matches 20; Conservative 10; Mismatches 19; Indels 3; Gaps 2;

QY 1 QRATIRGDAPGAGVPGGAVPGGAVPGGFLDYG-WYGVDSSTVDLAQ 51

Db 416 RRLLYSDAP--GSLGNISVPSTGATHVFCGLVPGAHYRVDIASSTGDISQ 465

RESULT 12

HEMA_IAME2

ID HEMA_IAME2 STANDARD; PRT; 566 AA.

AC P03439;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain].

GN HA.

OS Influenza A virus (strain A/Memphis/102/72).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=11439;

ON [1]

RP SEQUENCE FROM N.A.

RA Sleigh M.J., Both G.W., Brownlee G.G., Bender V.J., Moss B.A.;

RT "The haemagglutinin gene of influenza A virus: nucleotide sequence

RT analysis of cloned DNA copies.;"

RL (in) Laver G., Air G. (eds.);

RL Structure and variation in influenza virus, pp.69-79, Elsevier,

RL New York (1980).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.

CC -----

CC EMBL: V01089; CAA24273.1; -

CC PIR: A94441; HMTVHM.

CC HSSP: P03437; 2YIU.

CC InterPro: IPR001364; Hemagglutn.

CC Pfam: PF00509; Hemagglutinin; 1.

CC PRINTS: PR00329; HEMAGGLUTN12.

CC ProDom: PD00225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.

```

FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63264 MW; 1D9313AB3C380CD7 CRC64;

Query Match 23.7%; Score 64.5; DB 1; Length 566;
Best Local Similarity 20.2%; Pred. No. 17;
Matches 18; Conservative 10; Mismatches 10; Indels 51; Gaps 3;

QY 2 RATIRGDPAGAV-----PGAVPG----- 23
DB 280 KSSIMRSDAPICISICITPNSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGMRNV 339
QY 24 -----GAVPG-----GFGPLLDGWG 39
DB 340 PEKRTGLFGAIGAGFIENGWEGNIDGWG 368

RESULT 13
HS7A_CAEEL STANDARD; PRT; 640 AA.
AC P09446:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Heat shock 70 kDa protein A.
GN HSP-1 OR HSP70A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88297155; PubMed=28411196;
RA "Snutch T.P., Heschl M.F.P., Baillie D.L.;
RT "The Caenorhabditis elegans hsp70 gene family: a molecular genetic
RT characterization."
RL Gene 64:241-255(1988).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M18540; AAA28078.1;
CC PIR; JTO285; HKRW7A.
CC HSSP; P19120; 1HPM.
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70.1.
CC PRINTS; PR00301; HEATSHOCK70.
CC PRODOM; PD000089; Hsp70.1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC Heat shock; ATP-binding; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC FT CONFLICT 175 175 A -> G (IN REF. 2).
CC SQ SEQUENCE 640 AA; 69823 MW; ED15409D06C500C2 CRC64;

Query Match 23.7%; Score 64.5; DB 1; Length 640;
Best Local Similarity 58.3%; Pred. No. 19;
Matches 14; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 12 AGGAVPGGAVPGGAVGFGPLLD 35
DB 615 AGGA-PFGAIPGGAAGAGGTIE 637

```

```

RESULT 14
HS7L_CANAL STANDARD; PRT; 655 AA.
ID HS7L_CANAL
AC P41797;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Heat shock protein SSAL.
GN SSAL OR HSP70.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 20955;
RA Cassone A.C., la Valle R.L., Crisanti A.C., Muller H.M., Bromuro C.B.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-243 FROM N.A.
RX MEDLINE=95262895; PubMed=7742444;
RA Eroles P., Sentandreu M., Elorza M.V., Sentandreu R.;
RT "Cloning of a DNA fragment encoding part of a 70-kDa heat shock
RT protein of Candida albicans."
RL FEMS Microbiol. Lett. 128:95-100(1995).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z30210; CAA82929.1;
CC DR EMBL; S78163; AAB34280.1;
CC DR HSSP; P19120; 3HSC.
CC DR COMPLEVEAST-2DPAGE; P41797;
CC DR InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC PRODOM; PD000089; Hsp70.1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC Heat shock; ATP-binding; Multigene family; Acetylation.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC FT CONFLICT 175 175 A -> G (IN REF. 2).
CC SQ SEQUENCE 655 AA; 70192 MW; 634743E4D6DAD9D5 CRC64;

Query Match 23.7%; Score 64.5; DB 1; Length 655;
Best Local Similarity 48.8%; Pred. No. 19;
Matches 21; Conservative 2; Mismatches 15; Indels 5; Gaps 4;

QY 8 GDAPAG-GAVPG-GAVPGGAVPGGFGPLLDGWYGVYDVSSTVD 48
DB 613 GGAPGGAGGFGGAGGFGGGA-PGAGGP--GGATGESSGPTVE 652

RESULT 15
ELS-RAT STANDARD; PRT; 864 AA.
ID ELS-RAT
AC Q99372;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elastin precursor (Tropoelastin) (Fragment).
GN ELN.
OS Rattus norvegicus (Rat).

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 26.4991 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-7

Perfect score: 272

Sequence: 1 QRATRRGDAPAGGAVPGGA.....PLLDGNGVDVSDSTVDLAQ 51

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:**
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:**
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:**
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:**
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:**
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:**
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:**
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:**
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:**
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:**
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:**
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:**
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:**
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:**
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:**
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:**
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:**
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:**
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:**
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:**
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:**
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:**
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	272	100.0	51	23	AAE16186 B. bronchiseptica
2	272	100.0	922	13	AAE25578 Bordetella parapertussidis
3	272	100.0	922	13	AAE16185 Bordetella parapertussidis
4	263	96.7	922	12	AAE14321 Pertactin antigen
5	241.5	88.8	56	23	AAE16189 B. bronchiseptica
6	241.5	88.8	910	23	AAE16184 Bordetella pertussis
7	241.5	88.8	910	23	AAE17146 Bordetella pertussis
8	241.5	88.6	51	23	AAE16192 B. bronchiseptica
9	232.5	85.5	46	23	AAE16187 B. bronchiseptica
10	232.5	85.5	911	12	AAE14320 Pertactin antigen

11	232.5	85.5	911	13	AAE26503 prn proteins. Bor
12	232.5	85.5	911	23	AAE16183 Bordetella bronch
13	228.5	84.0	56	23	AAE16188 B. bronchiseptica
14	228.5	84.0	56	23	AAE16191 B. bronchiseptica
15	226	83.1	61	23	AAE16190 B. bronchiseptica
16	204.5	75.2	45	23	AAE17153 F647 monoclonal an
17	204.5	75.2	45	23	AAE17154 pEM3 monoclonal an
18	204.5	75.2	45	23	AAE17155 pEM4 monoclonal an
19	204.5	75.2	45	23	AAE17156 pEM68 monoclonal a
20	204.5	75.2	45	23	AAE17157 pEM70 monoclonal a
21	204.5	75.2	45	23	AAE17158 pEM71 monoclonal a
22	204.5	75.2	45	23	AAE17159 pEM72 monoclonal a
23	160.5	59.0	59	23	AAE17147 Bordetella pertuss
24	159.5	58.6	36	23	AAE17141 Bordetella pertuss
25	159.5	58.6	36	23	AAE17170 Maltose binding pr
26	155.5	57.2	56	23	AAE17152 Bordetella pertuss
27	151	55.5	54	23	AAE17150 Bordetella pertuss
28	150	55.1	31	23	AAE17173 Maltose binding pr
29	147.5	54.2	59	23	AAE17149 Bordetella pertuss
30	147.5	54.2	64	23	AAE17148 Bordetella pertuss
31	146.5	53.9	36	23	AAE17172 Maltose binding pr
32	146.5	53.9	41	23	AAE17171 Maltose binding pr
33	142.5	52.4	36	23	AAE17174 Maltose binding pr
34	138	50.7	54	23	AAE17151 Bordetella pertuss
35	99	36.4	20	23	AAE17144 Bordetella pertuss
36	86	31.6	15	23	AAE17166 Monoclonal Ab deri
37	85	31.2	15	23	AAE17163 Monoclonal Ab deri
38	81	29.8	15	23	AAE17161 Monoclonal Ab deri
39	81	29.8	15	23	AAE17162 Monoclonal Ab deri
40	81	29.8	398	22	ABE63199 Drosophila melanog
41	81	29.8	438	21	AAE17460 Arabidopsis thalia
42	81	29.8	460	21	AAE17459 Arabidopsis thalia
43	81	29.8	646	21	AAE17458 Arabidopsis thalia
44	77	28.3	15	23	AAE17160 Monoclonal Ab deri
45	77	28.3	450	23	AAU77399 Elastin-like pepti

ALIGNMENTS

RESULT 1

AAE16186

ID AAE16186 standard; peptide; 51 AA.

XX AAE16186;

AC AAE16186;

XX AAE16186;

XX AAE16186;

DT 26-MAR-2002 (first entry)

XX B. bronchiseptica strain I-1 pertactin outer membrane protein region I.

XX Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;

XX therapy; antibiotic; antibacterial; region I.

XX Bordetella bronchiseptica.

OS

XX WO200190143-A2.

PN 29-NOV-2001.

XX 23-MAY-2001; 2001WO-EP06457.

XX 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of

PT pertactin in Bordetella species, useful in immunogenic compositions for

PT treating infections caused by Bordetella and in diagnostic methods

XX

PS Claim 26; Fig 1b; 47pp; English.

XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region I.

XX Sequence 51 AA;
SQ

Query Match 100.0%; Score 272; DB 23; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.3e-24;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFLLDGWTGVDVSDSTVDLAQ 51
Db 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFLLDGWTGVDVSDSTVDLAQ 51

RESULT 2
AAR25578
ID AAR25578 standard; Protein; 922 AA.
XX
AC AAR25578;
XX
DT 08-JAN-1993 (first entry)
XX
DE Bordetella parapertussis P95 antigen precursor.
XX
KW Whooping cough; P70 antigen; P95 precursor protein; vaccination.

OS Bordetella parapertussis.
XX
FH Key Location/Qualifiers
FT Protein 35..643
FT /label- P70
FT Binding-site 260..262
FT /note= "motif associated with cell-cell adhesion"
FT Region 266..285
FT /note= "contains 5 direct, tandem repeats"
FT Region 575..612
FT /note= "contains 9 direct repeats of Pro-Gln-Pro"
FT Binding-site 712..714
FT /note= "motif associated with cell-cell adhesion"

XX W09211292-A.
XX
PD 09-JUL-1992.
XX
PF 23-DEC-1991; 91WO-GB02302.
XX
PR 21-DEC-1990; 90GB-0027901.
XX
PA (WELL) WELLCOME FOUND LTD.
XX
XX Charles IG;
XX WPI; 1992-250033/30.
XX N-PSDB; AAQ26509.
XX
XX Acellular vaccine for immunisation against whooping cough -
PT comprises protein uncontaminated by B. parapertussis components
PT and capable of binding antibodies which bind native P70 antigen
XX
PS Claim 1; Fig 1; 20pp; English.

XX A cosmid library was constructed by transforming E.coli HB101 with
CC recombinant cosmids prepared by partial digestion of B.parapertussis
CC chromosomal DNA with Sau3A and cloning of 40-50kb fragments into the
CC BamHI site of cosmid pHC79. The cosmids were screened with a 1.8kb
CC ClaI fragment from the prn gene of B.pertussis. The insert from one
CC positive colony, harbouring cosmid pBD811, was sequenced and found to
CC contain an open reading frame encoding a 922 amino acid protein
CC with calculated mol.wt. 95,177. This precursor protein ("P95") is
CC processed in vivo to the P70 antigen of apparent mol. wt. 70,000 as
CC determined by SDS-PAGE, but with actual mol.wt. 61kD. Antigenic
CC fragments of the protein will be useful in developing an acellular
CC vaccine against B.parapertussis. Preferred fragments include amino
CC acids Pro577 to Pro612 or Ala574 to Pro612.

XX Sequence 922 AA;
SQ

Query Match 100.0%; Score 272; DB 13; Length 922;
Best Local Similarity 100.0%; Pred. No. 4.2e-23;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFLLDGWTGVDVSDSTVDLAQ 51
Db 254 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFLLDGWTGVDVSDSTVDLAQ 304

RESULT 3
AAE16185
ID AAE16185 standard; Protein; 922 AA.
XX
AC AAE16185;
XX
DT 26-MAR-2002 (first entry)
XX
DE Bordetella parapertussis pertactin outer membrane protein, p.70.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; p.70.

OS Bordetella parapertussis.
XX
FH Key Location/Qualifiers
FT Region 254..304
FT /note= "Pertactin region I"
FT Region 564..621
FT /note= "Pertactin region II"
XX W0200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
XX WPI; 2002-097639/13.
XX N-PSDB; AAD26442.
XX
XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods -
XX
XX Disclosure; Page 34; 47pp; English.

XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when

CC antigens can be used for qualitative or quantitative determination of
 CC Bordetella in a material. The present sequence is B. bronchiseptica
 CC pertactin outer membrane protein region I.

SQ Sequence 56 AA;
 Query Match 88.8%; Score 241.5; DB 23; Length 56;
 Best Local Similarity 83.9%; Pred. No. 8.2e-21;
 Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWGVDSVSDTVDLAQ 51
 |||||
 Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGGFGPLLDGWGVDSVSDTVDLAQ 56

RESULT 6
 AAEL16184
 ID AAEL16184 standard; Protein; 910 AA.
 XX
 AC AAEL16184;
 DT 26-MAR-2002 (first entry)
 DE Bordetella pertussis pertactin outer membrane protein, p.69.
 KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
 KW therapy; antibiotic; antibacterial; p.69.
 OS Bordetella pertussis.

XX Key Location/Qualifiers
 FH Region 254..309
 FT /note= "Pertactin region I"
 FT Region 568..609
 FT /note= "Pertactin region II"
 XX WO200190143-A2.
 XX
 XX 29-NOV-2001.
 XX
 XX 23-MAY-2001; 2001WO-EP06457.
 XX
 XX 25-MAY-2000; 2000US-206969P.
 XX
 XX (INSP) INST PASTEUR.
 XX
 XX Guiso-maclouf N, Boursaux-eude C;
 XX
 XX WPI; 2002-097639/13.
 XX
 XX DR N-PSDB; AAD26441.

XX Polypeptides containing polymorphisms of the repeated regions of
 PT pertactin in Bordetella species, useful in immunogenic compositions for
 PT treating infections caused by Bordetella and in diagnostic methods
 XX
 XX Disclosure; Page 31; 47pp; English.

XX The present invention relates to Bordetella bronchiseptica pertactin
 CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
 CC vaccine. Pertactin antibody is useful for treating Bordetella infections
 CC and used to detect Bordetella antigens in biological preparations or in
 CC purifying corresponding proteins, glycoproteins or their mixtures when
 CC used in affinity chromatographic columns. Pertactin is useful as antigens
 CC to identify antibodies to Bordetella in materials such as human or other
 CC animal tissue and human or other animal cells, as well as biological
 CC fluids, such as human or other animal body fluids, including human sera,
 CC and to determine the concentration of Ab in those materials. Thus the
 CC antigens can be used for qualitative or quantitative determination of
 CC Bordetella in a material. The present sequence is B. pertussis
 CC pertactin outer membrane protein, p.69.

XX Sequence 910 AA;

Query Match 88.8%; Score 241.5; DB 23; Length 910;
 Best Local Similarity 83.9%; Pred. No. 1.4e-19;
 Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWGVDSVSDTVDLAQ 51
 |||||
 Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGGFGPLLDGWGVDSVSDTVDLAQ 309

RESULT 7
 AAEL17146
 ID AAEL17146 standard; Protein; 910 AA.
 XX
 AC AAEL17146;
 DT 18-APR-2002 (first entry)
 DE Bordetella pertussis pertactin (Prn1) protein.
 KW Pertactin; prn1; antibacterial; immunostimulant; antimicrobial; vaccine;
 KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
 KW immune response.

XX Bordetella pertussis.

XX Key Location/Qualifiers
 FH Region 597..604
 FT /note= "Conserved region"

XX WO200200695-A2.
 XX
 XX 03-JAN-2002.
 XX
 XX 29-JUN-2001; 2001WO-NL00493.
 XX
 XX 30-JUN-2000; 2000EP-0202309.
 XX
 XX (NEW-) NEDERLANDEN MIN WELZIJN.
 XX
 XX Mooi FR;
 XX
 XX WPI; 2002-139897/18.

XX New polypeptides derived from Bordetella pertussis pertactin, useful as
 a vaccine against infections caused by Bordetella strains, and other
 XX infectious diseases of mammals, e.g. diphtheria, tetanus, or polio
 XX
 XX Claim 11; Page 35-38; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
 CC pertactin (Prn1). The polypeptide is useful in the preparation of
 CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
 CC other infectious diseases of mammals including diphtheria, tetanus, polio
 CC and infections caused by Haemophilus influenza B. The polypeptide is
 CC especially useful for eliciting an immune response against Bordetella sp.
 CC Antibodies against the polypeptide may be used for pharmaceutical and/or
 CC diagnostic purposes, particularly for treating or preventing infections
 CC caused by Bordetella pertussis or Bordetella parapertussis. The present
 CC sequence is B. pertussis prn1 protein.

XX Sequence 910 AA;

Query Match 88.8%; Score 241.5; DB 23; Length 910;
 Best Local Similarity 83.9%; Pred. No. 1.4e-19;
 Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWGVDSVSDTVDLAQ 51
 |||||
 Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGGFGPLLDGWGVDSVSDTVDLAQ 309

RESULT 8
 AAEL16192

ID AAE16192 standard; peptide; 51 AA.

XX AC AAE16192;

XX DT 26-MAR-2002 (first entry)

XX DE B. bronchiseptica strain prn4 pertactin outer membrane protein region I.

XX DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;

XX KW therapy; antibiotic; antibacterial; region I.

XX OS Bordetella bronchiseptica.

XX PN WO200190143-A2.

XX PD 29-NOV-2001.

XX PF 23-MAY-2001; 2001WO-EP06457.

XX PR 25-MAY-2000; 2000US-206969P.

XX PA (INSP) INST PASTEUR.

XX PI Guiso-maclouf N, Boursaux-eude C;

XX DR WPI; 2002-097639/13.

XX PT Polypeptides containing polymorphisms of the repeated regions of
pertactin in Bordetella species, useful in immunogenic compositions for
treating infections caused by Bordetella and in diagnostic methods

XX PS Disclosure; Fig 1b; 47pp; English.

XX CC The present invention relates to Bordetella bronchiseptica pertactin
(outer membrane protein) or their fragments. Pertactin (PRN) is used as
vaccine. Pertactin antibody is useful for treating Bordetella infections
and used to detect Bordetella antigens in biological preparations or in
purifying corresponding proteins, glycoproteins or their mixtures when
used in affinity chromatographic columns. Pertactin is useful as antigens
to identify antibodies to Bordetella in materials such as human or other
animal tissue and human or other animal cells, as well as biological
fluids, such as human or other animal body fluids, including human sera,
and to determine the concentration of Ab in those materials. Thus the
antigens can be used for qualitative or quantitative determination of
Bordetella in a material. The present sequence is B. bronchiseptica
pertactin outer membrane protein region I.

XX SQ Sequence 51 AA;

Query Match 88.6%; Score 241; DB 23; Length 51;
Best Local Similarity 88.2%; Pred. No. 8.6e-21;
Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 QRATIRGDAPAGVPGGAVPGGPGPLLDGWYGVDSSTVDLAQ 51

|||||
Db 1 QRATIRGDAPAGVPGGAVPGGPGPLLDGWYGVDSSTVDLAQ 51

RESULT 9

AAE16187

ID AAE16187 standard; peptide; 46 AA.

XX AC AAE16187;

XX DT 26-MAR-2002 (first entry)

XX DE B. bronchiseptica strain I-2 pertactin outer membrane protein region I.

XX DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;

XX KW therapy; antibiotic; antibacterial; region I.

XX OS Bordetella bronchiseptica.

XX

PN WO200190143-A2.

XX PD 29-NOV-2001.

XX PF 23-MAY-2001; 2001WO-EP06457.

XX PR 25-MAY-2000; 2000US-206969P.

XX PA (INSP) INST PASTEUR.

XX PI Guiso-maclouf N, Boursaux-eude C;

XX DR WPI; 2002-097639/13.

XX PT Polypeptides containing polymorphisms of the repeated regions of
pertactin in Bordetella species, useful in immunogenic compositions for
treating infections caused by Bordetella and in diagnostic methods

XX PS Claim 26; Fig 1b; 47pp; English.

XX CC The present invention relates to Bordetella bronchiseptica pertactin
(outer membrane protein) or their fragments. Pertactin (PRN) is used as
vaccine. Pertactin antibody is useful for treating Bordetella infections
and used to detect Bordetella antigens in biological preparations or in
purifying corresponding proteins, glycoproteins or their mixtures when
used in affinity chromatographic columns. Pertactin is useful as antigens
to identify antibodies to Bordetella in materials such as human or other
animal tissue and human or other animal cells, as well as biological
fluids, such as human or other animal body fluids, including human sera,
and to determine the concentration of Ab in those materials. Thus the
antigens can be used for qualitative or quantitative determination of
Bordetella in a material. The present sequence is B. bronchiseptica
pertactin outer membrane protein region I.

XX SQ Sequence 46 AA;

Query Match 85.5%; Score 232.5; DB 23; Length 46;
Best Local Similarity 90.2%; Pred. No. 7.4e-20;
Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

OY 1 QRATIRGDAPAGVPGGAVPGGPGPLLDGWYGVDSSTVDLAQ 51

|||||
Db 1 QRATIRGDAPAGVPGGAVPGGPGPLLDGWYGVDSSTVDLAQ 46

RESULT 10

AAR14320

ID AAR14320 standard; Protein; 911 AA.

XX AC AAR14320;

XX DT 20-JAN-1992 (first entry)

XX DE Pertactin antigen P.68.

XX KW Pertactin; Pichia; B. pertussis; B. paraptussis.

XX OS Bordetella bronchiseptica.

XX FH Key Location/Qualifiers

FT Peptide 266..270

FT Peptide /label= repeat

FT Peptide 271..275

FT Peptide /label= repeat

FT Peptide 570..572

FT Peptide /label= repeat

FT Peptide 574..576

FT Peptide /label= repeat

FT Peptide 578..580

FT Peptide /label= repeat

FT Peptide 581..583

FT Peptide /label= repeat

FT Peptide 584..586

FT Peptide /label= repeat
 FT 587..589
 FT /label= repeat
 FT 599..601
 FT /label= repeat
 XX
 PN W09115571-A.
 XX
 XX 17-OCT-1991.
 XX
 XX 28-MAR-1991; 91WO-GB00487.
 XX
 XX 02-APR-1990; 90GB-0007416.
 XX
 XX (WELL) WELLCOME FOUNDATION LTD.
 XX
 XX Clare JJ, Romanos MA;
 XX
 XX WPI; 1991-325214/44.
 DR N-PSDB; AAQ14319.
 XX
 XX Pichia microorganism transformants - for production of
 PT Bordetella pertactin antigens for whooping cough vaccines
 XX
 PS Disclosure; Fig 1b; 38pp; English.
 XX
 XX Pichia microorganisms are transformed for the expression of
 CC pertactin antigens. DNA sequence used are represented in AAQ14319-20
 CC encoding the B. bronchiseptica P.68 and B. paraptussis P.70 antigen
 CC respectively or the B. pertussis P.69 encoding sequence described
 CC by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
 CC (1989).
 XX
 XX Sequence 911 AA;
 SQ
 Query Match 85.5%; Score 232.5; DB 12; Length 911;
 Best Local Similarity 90.2%; Pred. No. 1.5e-18;
 Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
 OY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 51
 DB 254 QRATIRGDAPA-----GGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 299
 RESULT 11
 AAR26503
 ID AAR26503 standard; Protein; 911 AA.
 XX
 AC AAR26503;
 XX
 XX 12-MAR-1993 (first entry)
 DT prn proteins.
 DE
 XX
 XX B. bronchiseptica; P.68; outer membrane protein; piglet; probe;
 KW atrophic rhinitis; alternative cleavage.
 XX
 OS Bordetella bronchiseptica.
 XX
 XX Key Location/Qualifiers
 FT Protein 35..632
 FT /label= P.68
 FT Region 266..279
 FT /label= Repeat_region
 FT Region 570..589
 FT /label= Repeat_region
 FT Peptide 260..262
 FT /label= RGD-tripeptide
 FT Peptide 701..703
 FT /label= RGD-tripeptide
 XX
 XX W09217587-A.
 PN
 XX

PD 15-OCT-1992.
 XX
 XX 27-MAR-1992; 92WO-GB00561.
 PF
 XX 27-MAR-1991; 91GB-0006568.
 PR
 XX (WELL) WELLCOME FOUND LTD.
 XX
 XX Charles IG;
 PI
 XX WPI; 1992-366258/44.
 DR N-PSDB; AAQ34566.
 XX
 XX DNA encoding a Bordetella bronchiseptica protein - used for
 PT obtaining vaccines for preventing respiratory diseases, partic.
 PT atrophic rhinitis in pigs
 XX
 PS Claim 1; Fig 1; 28pp; English.
 XX
 XX The sequence given is the P.94 antigen from B. bronchiseptica. The
 CC P.68 antigen is formed by alternative cleavage of this protein.
 CC P.68 is an outer membrane protein with a molecular weight of 68 kD
 CC which is associated with protection of piglets against atrophic
 CC rhinitis. The DNA sequence encoding these proteins was derived by
 CC standard recombinant DNA techniques using P.68 probes to isolate the
 CC entire P.94 sequence.
 XX
 SQ Sequence 911 AA;
 Query Match 85.5%; Score 232.5; DB 13; Length 911;
 Best Local Similarity 90.2%; Pred. No. 1.5e-18;
 Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
 OY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 51
 DB 254 QRATIRGDAPA-----GGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 299
 RESULT 12
 AAE16183
 ID AAE16183 standard; Protein; 911 AA.
 XX
 AC AAE16183;
 XX
 XX 26-MAR-2002 (first entry)
 DT Bordetella bronchiseptica pertactin outer membrane protein, p.68.
 XX
 XX Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
 KW therapy; antibiotic; antibacterial; p.68.
 XX
 OS Bordetella bronchiseptica.
 XX
 XX Key Location/Qualifiers
 FT Region 254..299
 FT /note= "Pertactin region I"
 FT Region 559..610
 FT /note= "Pertactin region II"
 XX
 XX WO200190143-A2.
 PN
 XX 29-NOV-2001.
 PD
 XX 23-MAY-2001; 2001WO-EP06457.
 PF
 XX 25-MAY-2000; 2000US-206969P.
 PR
 XX (INSP) INST PASTEUR.
 PA
 XX Guiso-maclouf N, Boursaux-eude C;
 PI
 XX WPI; 2002-097639/13.
 DR N-PSDB; AAD26440.
 DR

XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
XX Disclosure; Page 28; 47pp; English.
XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein, p.68.
XX
XX Sequence 911 AA;
XX
XX Query Match 85.5%; Score 232.5; DB 23; Length 911;
XX Best Local Similarity 90.2%; Pred. No. 1.5e-18;
XX Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
XX
XX QY 1 QRATIRRGDAPAGGAVPGGAVPGGPGPLLDGMYGVDVSDSTVDLAQ 51
XX |||||
XX DB 254 QRATIRRGDAPAGGAVPGGAVPGGPGPLLDGMYGVDVSDSTVDLAQ 299
XX |||||
XX
XX RESULT 13
XX AAEL16188
XX ID AAEL16188 standard; peptide; 56 AA.
XX AC AAEL16188;
XX XX
XX DT 26-MAR-2002 (first entry)
XX DE B. bronchiseptica strain I-3 pertactin outer membrane protein region I.
XX KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX KW therapy; antibiotic; antibacterial; region I.
XX OS Bordetella bronchiseptica.
XX XX
XX PN WO200190143-A2.
XX XX
XX PD 29-NOV-2001.
XX XX
XX PF 23-MAY-2001; 2001WO-EP06457.
XX XX
XX PR 25-MAY-2000; 2000US-206969P.
XX XX
XX PA (INSP) INST PASTEUR.
XX XX
XX PI Guiso-maclouf N, Boursaux-eude C;
XX XX
XX DR WO200190143-A2.
XX XX
XX PD 29-NOV-2001.
XX XX
XX PF 23-MAY-2001; 2001WO-EP06457.
XX XX
XX PR 25-MAY-2000; 2000US-206969P.
XX XX
XX PA (INSP) INST PASTEUR.
XX XX
XX PI Guiso-maclouf N, Boursaux-eude C;
XX XX
XX DR WPI; 2002-097639/13.
XX XX
XX PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
XX Claim 26; Fig 1b; 47pp; English.
XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens

CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region I.
XX
XX Sequence 56 AA;
XX
XX Query Match 84.0%; Score 228.5; DB 23; Length 56;
XX Best Local Similarity 80.4%; Pred. No. 2.6e-19;
XX Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;
XX
XX QY 1 QRATIRRGDAPAGGAVPGGAVPGGPGPLLDGMYGVDVSDSTVDLAQ 51
XX |||||
XX DB 1 QRATIRRGDAPAGGAVPGGAVPGGPGPLLDGMYGVDVSDSTVDLAQ 56
XX |||||
XX
XX RESULT 14
XX AAEL16191
XX ID AAEL16191 standard; peptide; 56 AA.
XX AC AAEL16191;
XX XX
XX DT 26-MAR-2002 (first entry)
XX DE B. bronchiseptica strain prn3 pertactin outer membrane protein region I.
XX KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX KW therapy; antibiotic; antibacterial; region I.
XX OS Bordetella bronchiseptica.
XX XX
XX PN WO200190143-A2.
XX XX
XX PD 29-NOV-2001.
XX XX
XX PF 23-MAY-2001; 2001WO-EP06457.
XX XX
XX PR 25-MAY-2000; 2000US-206969P.
XX XX
XX PA (INSP) INST PASTEUR.
XX XX
XX PI Guiso-maclouf N, Boursaux-eude C;
XX XX
XX DR WPI; 2002-097639/13.
XX XX
XX PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
XX Disclosure; Fig 1b; 47pp; English.
XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region I.
XX
XX Sequence 56 AA;
XX
XX Query Match 84.0%; Score 228.5; DB 23; Length 56;
XX Best Local Similarity 80.4%; Pred. No. 2.6e-19;
XX Matches 45; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAV-----PGGAVPGGFGPLDGGWYGVDSSTVDLAQ 51
 Db |||||
 1 QRATIRRGDAPAGGAVPGGAVPGGFGPGGFGPLDGGWYGVDSSTVDLAQ 56

RESULT 15

AAE16190
 ID AAE16190 standard; peptide; 61 AA.
 XX
 AC AAE16190;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE B. bronchiseptica strain prn2 pertactin outer membrane protein region I.
 XX
 KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
 KW therapy; antibiotic; antibacterial; region I.
 XX
 OS Bordetella bronchiseptica.
 XX
 PN WO200190143-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 23-MAY-2001; 2001WO-EP06457.
 XX
 PR 25-MAY-2000; 2000US-206969P.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Guiso-maclouf N, Boursaux-eude C;
 XX
 DR WPI; 2002-097639/13.
 XX
 PT Polypeptides containing polymorphisms of the repeated regions of
 PT pertactin in Bordetella species, useful in immunogenic compositions for
 PT treating infections caused by Bordetella and in diagnostic methods
 XX
 PS Disclosure; Fig 1b; 47pp; English.
 XX
 CC The present invention relates to Bordetella bronchiseptica pertactin
 CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
 CC vaccine. Pertactin antibody is useful for treating Bordetella infections
 CC and used to detect Bordetella antigens in biological preparations or in
 CC purifying corresponding proteins, glycoproteins or their mixtures when
 CC used in affinity chromatographic columns. Pertactin is useful as antigens
 CC to identify antibodies to Bordetella in materials such as human or other
 CC animal tissue and human or other animal cells, as well as biological
 CC fluids, such as human or other animal body fluids, including human sera,
 CC and to determine the concentration of Ab in those materials. Thus the
 CC antigens can be used for qualitative or quantitative determination of
 CC Bordetella in a material. The present sequence is B. bronchiseptica
 CC pertactin outer membrane protein region I.
 XX
 SQ Sequence 61 AA;

Query Match 83.1%; Score 226; DB 23; Length 61;
 Best Local Similarity 73.8%; Pred. No. 5.5e-19;
 Matches 45; Conservative 3; Mismatches 3; Indels 10; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAV-----PGGAVPGGFGPLDGGWYGVDSSTVDLA 50
 Db |||||
 1 QRATIRRGDAPAGGAVPGGAVPGGFGPGGFGPLDGGWYGVDSSTVDLA 60
 QY 51 Q 51
 Db 61 Q 61

Search completed: May 7, 2003, 16:47:02
 Job time : 37.4991 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:41:51 ; Search time 8.4276 Seconds
(without alignments)
178.054 Million cell updates/sec

Title: US-09-855-754B-7
Perfect score: 272
Sequence: 1 QRATIRGDAPAGVPGGAVP...PLLDGWGVDSSTVDLAQ 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/PTUS_COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272	100.0	922	4	US-08-460-269C-6
2	232.5	85.5	911	4	US-08-460-269C-4
3	231.5	85.1	910	4	US-08-460-269C-2
4	94.5	34.7	745	2	US-09-010-928B-28
5	94.5	34.7	870	2	US-09-010-928B-2
6	89	32.7	907	2	US-09-010-928B-4
7	76	27.9	20	2	US-09-010-928B-9
8	68	25.0	4545	2	US-08-804-227C-14
9	68	25.0	4550	2	US-08-804-227C-8
10	68	25.0	4550	2	US-08-804-198-2
11	67	24.6	420	2	US-08-845-998-8
12	67	24.6	420	4	US-09-206-537-8
13	67	24.6	420	4	US-09-430-854-8
14	66	24.3	20	2	US-09-010-928B-7
15	65.5	24.1	1711	2	US-08-342-930-2
16	64	23.5	119	4	US-08-556-978B-60
17	64	23.5	119	4	US-09-247-806-9
18	64	23.5	493	4	US-08-556-978B-59
19	64	23.5	529	1	US-09-247-806-2
20	64	23.5	595	1	US-08-425-069-4
21	64	23.5	595	2	US-08-317-844B-4
22	64	23.5	714	4	US-08-556-978B-61
23	64	23.5	714	4	US-09-247-806-10
24	63.5	23.3	262	3	US-08-946-914-14
25	63.5	23.3	262	4	US-09-656-450-14
26	63.5	23.3	566	4	US-09-232-468A-22
27	63	23.2	235	2	US-08-529-190B-1

28	63	23.2	426	1	US-07-918-023-2	Sequence 2, Appl
29	63	23.2	641	4	US-09-249-585A-3	Sequence 3, Appl
30	62.5	23.0	304	5	PCT-US94-01149-32	Sequence 32, Appl
31	62.5	23.0	571	2	US-08-453-848-15	Sequence 15, Appl
32	62.5	23.0	571	2	US-08-453-848-21	Sequence 21, Appl
33	62.5	23.0	571	4	US-09-169-027-15	Sequence 15, Appl
34	62.5	23.0	571	4	US-09-169-027-21	Sequence 21, Appl
35	62	22.8	60	4	US-08-924-629C-31	Sequence 31, Appl
36	62	22.8	264	1	US-08-562-311-4	Sequence 4, Appl
37	62	22.8	574	4	US-09-276-400-6	Sequence 6, Appl
38	62	22.8	574	4	US-09-448-076-6	Sequence 6, Appl
39	62	22.8	574	4	US-09-702-572-6	Sequence 6, Appl
40	61	22.4	43	3	US-08-542-051-31	Sequence 31, Appl
41	61	22.4	43	3	US-08-542-051-32	Sequence 32, Appl
42	61	22.4	43	3	US-08-542-051-33	Sequence 33, Appl
43	61	22.4	48	3	US-08-542-051-45	Sequence 45, Appl
44	61	22.4	160	3	US-08-542-051-18	Sequence 18, Appl
45	61	22.4	251	1	US-08-397-633A-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-08-460-269C-6
; Sequence 6, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6

Query Match 100.0%; Score 272; DB 4; Length 922;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGVPGGAVP...PLLDGWGVDSSTVDLAQ 51
DB 254 QRATIRGDAPAGVPGGAVP...PLLDGWGVDSSTVDLAQ 304

RESULT 3
 US-08-460-269C-2
 : Sequence 2, Application US/08460269C
 : Patent No. 6197548
 : GENERAL INFORMATION:
 : APPLICANT: CLARE, JEFFREY J.
 : ROMANOS, MICHAEL A.
 : TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
 : YEAST
 : NUMBER OF SEQUENCES: 17
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Millien, White, Zelano & Branigan, P.C.
 : STREET: 2200 Clarendon Blvd., Suite 1400
 : CITY: ARLINGTON
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22201
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:

```
RESULT 5
US-09-010-928B-2
; Sequence 2, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010, 928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-2

Query Match 34.7%; Score 94.5; DB 2; Length 870;
Best Local Similarity 65.6%; Pred. No. 0.0014;
Matches 21; Conservative 0; Mismatches 8; Indels 3; Gaps 1;

QY 8 GDAPAGGAVPGGAVPGGFGPLDGGWG 39
| | | | | | | | | | | | | | | | | | | | |
DB 369 GGAGPGGAGGGAGGGAGGGAGP---GGYG 397

RESULT 6
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010, 928B
; FILING DATE: 22-JAN-1998
```

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-4

Query Match 32.7%; Score 89; DB 2; Length 907;
Best Local Similarity 72.0%; Pred. No. 0.0067;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 8 GDAPAGGAVPGGAVPGGFGFGP 32
| | | | | | | | | | | | | | | | | | | | |
DB 163 GGAGPGGAGGGAGGGAGGGAGP 187

RESULT 7
US-09-010-928B-9
; Sequence 9, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010, 928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-9

Query Match 27.9%; Score 76; DB 2; Length 20;
Best Local Similarity 78.9%; Pred. No. 0.0035;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 13 GGAVPGGAVPGGAVPGGFG 31
| | | | | | | | | | | | | | | | | | | | |
DB 2 GGAGPGGAGGGAGGGAGGAG 20

RESULT 8
US-08-804-227C-14
; Sequence 14, Application US/08804227C
; Patent No. 5876991
```

GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-14

Query Match 25.0%; Score 68; DB 2; Length 4545;
Best Local Similarity 48.5%; Pred. No. 15;
Matches 16; Conservative 1; Mismatches 12; Indels 4; Gaps 1;

Qy 12 AGGAVPGGAVPGGAVPGGFGPLLDG----WYGV 40
Db 1571 AAGRVSGVRSRAVPGGVLFTGCGAQWVGM 1603

RESULT 9
US-08-804-227C-8
Sequence 8, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-8
Query Match 25.0%; Score 68; DB 2; Length 4550;
Best Local Similarity 48.5%; Pred. No. 15;
Matches 16; Conservative 1; Mismatches 12; Indels 4; Gaps 1;

Qy 12 AGGAVPGGAVPGGAVPGGFGPLLDG----WYGV 40
Db 1576 AAGRVSGVRSRAVPGGVLFTGCGAQWVGM 1608

RESULT 10
US-08-804-198-2
Sequence 2, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rosteck, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-198-2

Query Match 25.0%; Score 68; DB 2; Length 4550;
Best Local Similarity 48.5%; Pred. No. 15;
Matches 16; Conservative 1; Mismatches 12; Indels 4; Gaps 1;

Qy 12 AGGAVPGGAVPGGAVPGGFGPLLDG----WYGV 40
Db 1576 AAGRVSGVRSRAVPGGVLFTGCGAQWVGM 1608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,854
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/845,998
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	272	100.0	51	9	US-09-855-754-7		Sequence 7, Appl1
2	272	100.0	922	9	US-09-855-754-5		Sequence 6, Appl1
3	248.5	91.4	910	9	US-09-855-754-5		Sequence 5, Appl1
4	241.5	88.8	56	9	US-09-855-754-10		Sequence 10, Appl1
5	241	88.6	51	9	US-09-855-754-13		Sequence 13, Appl1
6	232.5	85.5	46	9	US-09-855-754-8		Sequence 8, Appl1
7	232.5	85.5	911	9	US-09-855-754-4		Sequence 4, Appl1
8	228.5	84.0	56	9	US-09-855-754-9		Sequence 9, Appl1
9	228.5	84.0	56	9	US-09-855-754-12		Sequence 12, Appl1
10	226	83.1	61	9	US-09-855-754-11		Sequence 11, Appl1
11	77	28.3	450	10	US-09-812-382-6		Sequence 6, Appl1
12	76	27.9	50	10	US-09-812-382-3		Sequence 3, Appl1
13	69.5	25.6	111	10	US-09-837-969A-60		Sequence 60, Appl1
14	69.5	25.6	111	10	US-09-841-321A-60		Sequence 60, Appl1
15	67	24.6	48	10	US-09-837-969A-9		Sequence 9, Appl1
16	67	24.6	48	10	US-09-841-321A-9		Sequence 9, Appl1
17	65.5	24.1	457	9	US-10-078-770-194		Sequence 194, App
18	64	23.5	119	10	US-09-861-597-9		Sequence 9, Appl1
19	64	23.5	529	10	US-09-861-597-2		Sequence 2, Appl1

TITLE OF INVENTION: POLYPEPTIDES C

APPLICANT: GUISSO-MACLOUF, NICOL
TITLE OF INVENTION: POYDRETTES

4. TITLE

```
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,  
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA  
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS  
; FILE REFERENCE: 03495-0206-00000  
; CURRENT APPLICATION NUMBER: US/09/855,754  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 922  
; TYPE: PRT  
; ORGANISM: Bordetella parapertussis  
US-09-855-754-6  
  
Query Match 100.0%; Score 272; DB 9; Length 922;  
Best Local Similarity 100.0%; Pred. No. 5.5e-21;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 51  
|||||  
Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 304  
|||||  
  
RESULT 3  
US-09-855-754-5  
; Sequence 5, Application US/09855754  
; Publication No. US20020192237A1  
; GENERAL INFORMATION:  
; APPLICANT: BOURSAX-EUDE, CAROLINE  
; APPLICANT: GUISSO-MACLOUF, NICOLE  
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED  
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,  
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA  
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS  
; FILE REFERENCE: 03495-0206-00000  
; CURRENT APPLICATION NUMBER: US/09/855,754  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 910  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
US-09-855-754-5  
  
Query Match 91.4%; Score 248.5; DB 9; Length 910;  
Best Local Similarity 85.7%; Pred. No. 1.6e-18;  
Matches 48; Conservative 3; Mismatches 0; Indels 5; Gaps 1;  
  
Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 51  
|||||  
Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 309  
|||||  
  
RESULT 4  
US-09-855-754-10  
; Sequence 10, Application US/09855754  
; Publication No. US20020192237A1  
; GENERAL INFORMATION:  
; APPLICANT: BOURSAX-EUDE, CAROLINE  
; APPLICANT: GUISSO-MACLOUF, NICOLE  
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED  
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,  
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA  
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS  
; FILE REFERENCE: 03495-0206-00000
```

```
; CURRENT APPLICATION NUMBER: US/09/855,754  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Bordetella bronchiseptica  
US-09-855-754-10  
  
Query Match 88.8%; Score 241.5; DB 9; Length 56;  
Best Local Similarity 83.9%; Pred. No. 5.7e-19;  
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;  
  
Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 51  
|||||  
Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 56  
|||||  
  
RESULT 5  
US-09-855-754-13  
; Sequence 13, Application US/09855754  
; Publication No. US20020192237A1  
; GENERAL INFORMATION:  
; APPLICANT: BOURSAX-EUDE, CAROLINE  
; APPLICANT: GUISSO-MACLOUF, NICOLE  
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED  
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,  
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA  
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS  
; FILE REFERENCE: 03495-0206-00000  
; CURRENT APPLICATION NUMBER: US/09/855,754  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Bordetella bronchiseptica  
US-09-855-754-13  
  
Query Match 88.6%; Score 241; DB 9; Length 51;  
Best Local Similarity 88.2%; Pred. No. 5.9e-19;  
Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 51  
|||||  
Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFPGLLDGWYGVDSSTVDLAQ 51  
|||||  
  
RESULT 6  
US-09-855-754-8  
; Sequence 8, Application US/09855754  
; Publication No. US20020192237A1  
; GENERAL INFORMATION:  
; APPLICANT: BOURSAX-EUDE, CAROLINE  
; APPLICANT: GUISSO-MACLOUF, NICOLE  
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED  
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,  
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA  
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS  
; FILE REFERENCE: 03495-0206-00000  
; CURRENT APPLICATION NUMBER: US/09/855,754  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/206,969  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 24
```



```

; GENERAL INFORMATION:
; APPLICANT: UITY, Dan
; TITLE OF INVENTION: Injectable Implants For Tissue Augmentation and Restoration
; FILE REFERENCE: BERL-020/030US
; CURRENT APPLICATION NUMBER: US/09/837,969A
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/258,723
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: US 60/087155
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: US 60/076297
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(48)
; OTHER INFORMATION: Synthetic
US-09-837-969A-9

```

```

Query Match      24.6%; Score 67; DB 10; Length 48;
Best Local Similarity 68.0%; Pred. No. 1.2;
Matches 17; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

```

```

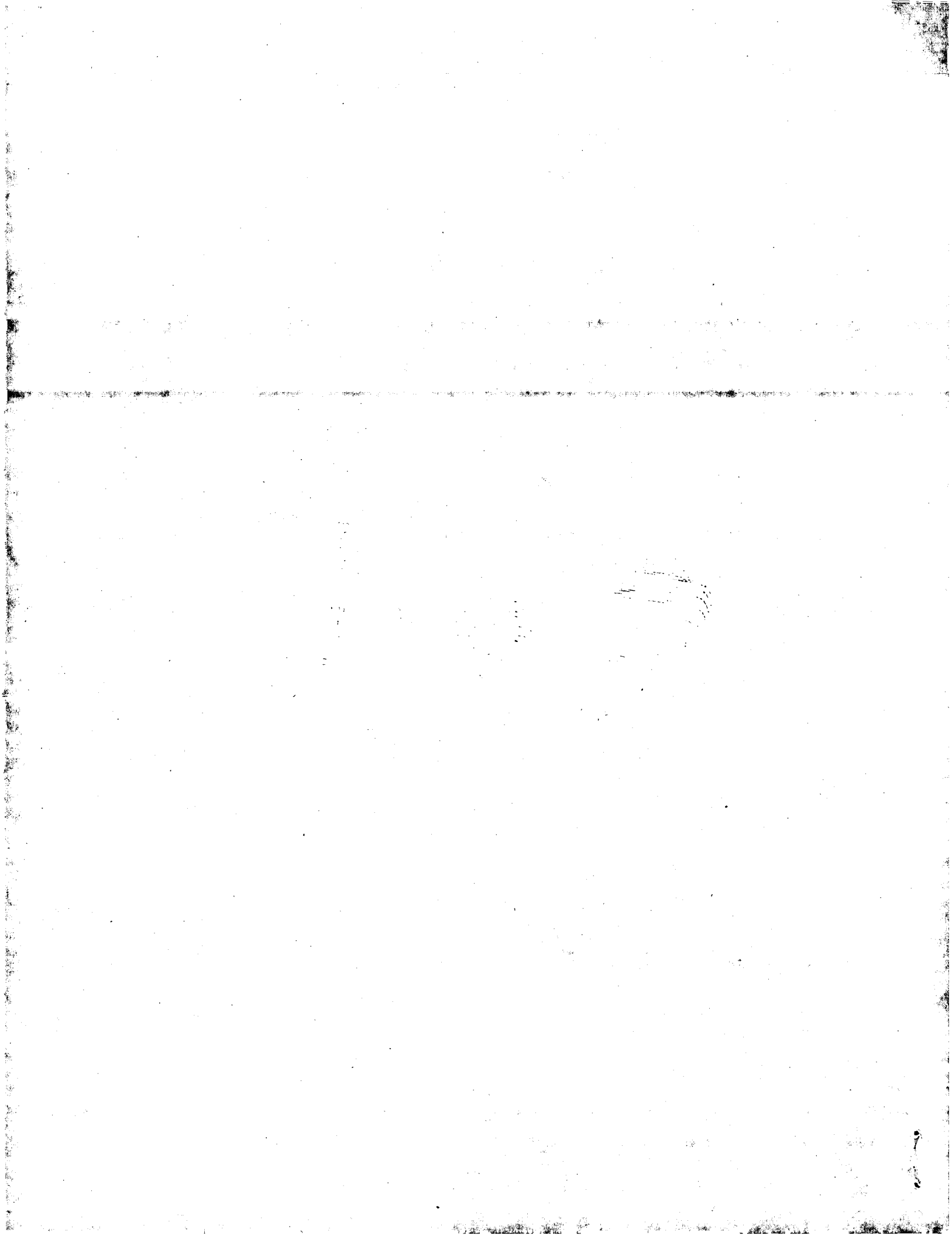
QY      8 GDAPAG--GAVPGGAVPGGAVPGGF 30
      | | | | | | | | | | | | | |
Db       9 GGVPGGFPGGVPGG-VPGG-VPGGF 31

```

```

Search completed: May 7, 2003, 17:30:47
Job time : 16.2913 secs

```



Result No.	Query			ID	Description
	Score	Match	Length		
1	272	100.0	922	S15204	peractin - Bordet
2	232.5	85.5	911	A47675	68K outer membrane
3	231.5	85.1	910	A32560	outer membrane pro
4	83	30.5	783	T35389	probable serine-th
5	81	29.8	646	T45677	ATP-dependent RNA
6	78	28.7	128	S25103	MFS18 protein - ma
7	74	27.2	860	EAMS	elastin precursor
8	70.5	25.9	680	A28121	major merozoite su
9	70.5	25.9	686	A49242	dnak-type molecula
10	70.5	25.9	1772	A45532	major merozoite su
11	69.5	25.6	641	PC7036	heat shock protein
12	69	25.4	112	J01063	glycine-rich prote
13	69	25.4	422	T24865	hypothetical prote
14	68.5	25.2	405	T29167	hypothetical prote
15	68.5	25.2	1300	T03166	probable immediate
16	68	25.0	179	A85217	hypothetical prote
17	68	25.0	277	T04441	hypothetical prote
18	67.5	24.8	231	S37108	cuticlin 2 - Caeno
19	67.5	24.8	907	A45560	sporozoite surface
20	67	24.6	76	T09262	glycine-rich cell
21	67	24.6	420	I59234	octamer binding tr
22	66.5	24.4	524	A63328	proteinase do (EC
23	66	24.3	361	T30743	hypothetical prote
24	65.5	24.1	76	I45885	elastin - bovine (
25	65.5	24.1	136	C95291	hypothetical prote
26	65.5	24.1	185	T49890	glycine-rich prote
27	65.5	24.1	1711	A55148	protein-tyrosine-p
28	65	23.9	310	B75476	conserved hypothet
29	65	23.9	514	B87592	hypothetical prote

RESULT 8

A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C:Accession: A28121
R:Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A:Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen en et
A:Reference number: A28121; MUID:88124869; PMID:2448778
A:Accession: A28121
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-680 <BUR>
A:Cross-references: GB:J03612; NID:gl60678; PID:gl60679
A:Experimental source: strain 17XL
A>Note: the authors translated the codon GTA for residue 429 as Leu
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 25.9%; Score 70.5; DB 2; Length 680;
Best Local Similarity 44.1%; Pred. No. 5.7;
Matches 26; Conservative 2; Mismatches 8; Indels 23; Gaps 5;

QY 4 TIRRGDAPA-----CGAVPG---GAVPG---GAVPGGFGPLLDGWGYVD--VSDSTVD 48
| | | | | | | | | | | | | | | | | | | | | | : | : | : |
DB 238 TTLAADAPATPEGAVPGAVPGAVPGAVPGAVPGAVPGAVPGVCS-----GTDTRVAGSSVD 287

RESULT 9

A49242
dnak-type molecular chaperone hsp70 - Plasmodium cynomolgi
N:Alternate names: heat shock protein hsp70
C:Species: Plasmodium cynomolgi
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A49242
R:Eckert, V.; Sanchez, L.; Cochran, A.; Enea, V.
Exp. Parasitol. 75, 323-328, 1992
A:Title: Plasmodium cynomolgi: the hsp 70 gene.
A:Reference number: A49242; MUID:93050041; PMID:1426134
A:Accession: A49242
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-686 <ECK>
A:Cross-references: GB:I90978; NID:gl60349; PIDN:AAA29625.1; PID:g160350
A>Note: sequence extracted from NCBI backbone (NCBIN:118975, NCBIP:118976)
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein compl
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 25.9%; Score 70.5; DB 2; Length 686;
Best Local Similarity 46.9%; Pred. No. 5.7;
Matches 15; Conservative 4; Mismatches 8; Indels 5; Gaps 2;

QY 8 GDAPAG---GAVPGGAVPGGAVPGGFGPLLD 35
| | | | | | | | | | | | | | | | | : | : | : | : | : |
DB 653 GGMPGGWNFPGGMPGGMPGGA-PAGSGPTVE 683

RESULT 10

A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C:Accession: A45532; A45531
R:Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A:Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A:Reference number: A45532; MUID:90014981; PMID:2797063
A:Accession: A45532
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1772 <LEW>

[illegible]

RESULT
T24865

Query Match 25.4%; Score 69; DB 2; Length 422;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 13; Conservative 1; Mismatches 10; Indels 2; Gaps 1;

RESULT 14

Query Match 25.2%; Score 68.5; DB 2; Length 405;
Best Local Similarity 41.0%; Pred. No. 5,6;
Matches 16; Conservative 6; Mismatches 16; Indels 1; Gaps 1;

RESULT 15

probable immediate early protein - alcelaphine herpesvirus 1
C;Species: alcelaphine herpesvirus 1

QY 8 GDAPAGGAVPGGA -VPGG - -AVPGSGFP 32
 | : | | | | | | | | | | | |
Dd 102 GEGPGGGEVPGGGEVPGGGGCGPGGEGP 129

Search completed: May 7, 2003, 16:55:19
Job time : 24.5571 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 4.9523 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-7

Perfect score: 272

Sequence: 1 QRATIRGDAPAGGAVPGGA.....PLLDGWYGVDSSTVDLAQ 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272	100.0	922	1	PERT_BORPA
2	241.5	88.8	910	1	PERT_BORPE
3	232.5	85.5	911	1	PERT_BORBR
4	78	28.7	128	1	MF18_MAIZE
5	74	27.2	860	1	ELS_MOUSE
6	70.5	25.9	586	1	HS70_PLACB
7	70.5	25.9	1772	1	MSPI_PLAYO
8	67.5	24.8	231	1	CUT2_CAEEL
9	67	24.6	423	1	BR3A_HUMAN
10	65.5	24.1	421	1	BR3A_MOUSE
11	65.5	24.1	1711	1	PTPO_RAT
12	64.5	23.7	566	1	HEMA_IAME2
13	64.5	23.7	640	1	HEMA_IAME2
14	64.5	23.7	655	1	HS71_CANAL
15	64.5	23.7	864	1	ELS_RAT
16	64	23.5	480	1	RADA_MYCTU
17	64	23.5	627	1	SPD2_NEPCL
18	63.5	23.3	261	1	LEGG_RAT
19	63.5	23.3	565	1	HEMA_IATKP
20	63.5	23.3	747	1	ELS_BOVIN
21	63.5	23.3	750	1	ELS_CHICK
22	63	23.2	295	1	LEGG_CANFA
23	63	23.2	342	1	ROAL_SCHAM
24	63	23.2	641	1	EBN1_EBV
25	63	23.2	778	1	YQ34_MYCTU
26	62.5	23.0	152	1	HS70_ALTAL
27	62.5	23.0	566	1	HEMA_TAUDO
28	62.5	23.0	644	1	HS71_PICAN
29	62.5	23.0	1174	1	KPCI_COCH
30	62	22.8	244	1	LEGG_CRILLO
31	62	22.8	263	1	LEGG_MOUSE
32	62	22.8	346	1	RX_HUMAN
33	62	22.8	574	1	LOLI_HUMAN

RESULT 1

ID	PERT_BORPA	STANDARD;	PRT;	922 AA.
AC	P24328;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.70) (P.95).			
GN	PRN.			
OS	Bordetella parapertussis.			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=519;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN2591;			
RX	MEDLINE=91251771; PubMed=2041476;			
RA	Li L.J., Dougan G., Novotny P., Charles I.G.;			
RT	"P.70 pertactin, an outer-membrane protein from Bordetella			
RT	parapertussis: cloning, nucleotide sequence and surface expression in			
RT	Escherichia coli."			
RL	Mol. Microbiol. 5:409-417(1991).			
CC	-1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS			
CC	MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN			
CC	BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SUBCELLULAR LOCATION: Outer membrane.			
CC	-1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)			
CC	CONCENTRATIONS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X54547; CAA38419.1; -			
DR	EMBL: A26124; CAA01786.1; -			
DR	EMBL: A19182; CAA01454.1; -			
DR	PIR: S15204; S15204.			
DR	PIR: S14659; S14659.			
DR	InterPro: IPR004899; Pertactin_sup.			
DR	InterPro: IPR003992; pertactin.			
DR	InterPro: IPR003991; pertactin_vir.			
DR	Pfam: PF03212; Pertactin; 1.			
DR	PRINTS; PR01482; PERTACTIN.			
DR	PRINTS; PR01484; PERTACTINFAMLY.			
KW	Outer membrane; Signal; Virulence; Repeat.			
FT	SIGNAL 1 34 POTENTIAL.			
FT	CHAIN 35 922 P.95.			
FT	CHAIN 35 647 PERTACTIN (P.70).			
FT	PROPEP 648 922 POTENTIAL.			
FT	SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN			

Q10265 schizosacch
P16770 human cytom
O35984 mus musculu
P23118 strongyloce
P11133 influenza a
P03440 influenza a
P26139 influenza a
P03435 influenza a
P10209 herpes simp
Q01877 puccinia gr
P47845 oryctolagus
Q10707 mycobacteri

ALIGNMENTS

```
FT FT ADHESION TO VARIOUS EUKARYOTIC CELL
FT FT LINES).
FT FT 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT FT 1.
FT FT 266 290
FT FT REPEAT 266 270
FT FT 2.
FT FT 271 275
FT FT 3.
FT FT 276 280
FT FT REPEAT 281 285
FT FT 4 (APPROXIMATE).
FT FT 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
FT FT 575 603
FT FT DOMAIN 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
SQ SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;

Query Match 100.0%; Score 272; DB 1; Length 922;
Best Local Similarity 100.0%; Pred. No. 7.3e-20;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGVPGGAVPGGFGPLLDGWGVDSSTVDLAQ 51
Db 254 QRATIRGDAPAGVPGGAVPGGFGPLLDGWGVDSSTVDLAQ 304

RESULT 2
ID PERT_BORPE STANDARD; PRT; 910 AA.
AC P14283;
DC 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CN2992;
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chaffield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis."
RT Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558 (1989).
RN [2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica."
RA J. Gen. Microbiol. 138:1697-1705 (1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RA Embley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin."
RT Nature 381:90-92 (1996).
CC CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC CC -1- SUBUNIT: MONOMER.
CC CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CC CONCENTRATION.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR PIR; A32560; A32560.
DR InterPro; IPR004899; Pertactin_sup.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMILY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 910 P.93
FT CHAIN 35 711 PERTACTIN (P.69).
FT PROPEP 712 910 POTENTIAL.
FT SITE 260 262 ADHESION TO VARIOUS EUKARYOTIC CELL
FT LINES).
FT FT 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT FT 1.
FT FT 266 290
FT FT REPEAT 266 270
FT FT 2.
FT FT 271 275
FT FT 3.
FT FT 276 280
FT FT REPEAT 281 285
FT FT REPEAT 286 290
FT FT DOMAIN 579 593
FT FT 5 X 3 AA TANDEM REPEATS OF P-Q-P.
SQ SEQUENCE 910 AA; 93452 MW; A169871E20A2E7DB CRC64;

Query Match 88.8%; Score 241.5; DB 1; Length 910;
Best Local Similarity 83.9%; Pred. No. 7.6e-17;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGVPGGAVPGGAVPGGFGPLLDGWGVDSSTVDLAQ 51
Db 254 QRATIRGDAPAGVPGGAVPGGFGPLLDGWGVDSSTVDLAQ 309

RESULT 3
ID PERT_BORBR STANDARD; PRT; 911 AA.
AC Q03035;
DC 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.68) (P.94).
GN PRN
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 35-44.
RC STRAIN=CN7531;
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica."
RA J. Gen. Microbiol. 138:1697-1705 (1992).
CC CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC CC -1- SUBUNIT: MONOMER.
CC CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CC CONCENTRATIONS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

EMBL; X54815; CAA38584.1; -

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	272	100.0	158	2	Q9K5G6		Q9k5g6 bordetella
2	272	100.0	159	2	Q9K5G8		Q9k5g8 bordetella
3	272	100.0	160	2	Q9K510		Q9k510 bordetella
4	272	100.0	179	2	Q9AHP1		Q9ahp1 bordetella
5	272	100.0	211	2	Q9KJY3		Q9k1y3 bordetella
6	259.5	95.4	216	2	Q9KJY4		Q9k1y4 bordetella
7	241.5	88.8	387	2	Q9S3M9		Q9s3m9 bordetella
8	241.5	88.8	910	2	Q9SGN1		Q9sgn1 bordetella
9	241	88.6	382	2	Q9J199		Q9j199 bordetella
10	241	88.6	905	2	Q9SGM9		Q9sgm9 bordetella
11	241	88.6	905	2	Q9SGN0		Q9sgn0 bordetella
12	239	87.5	167	2	Q937U5		Q937u5 bordetella
13	237.5	87.3	907	2	Q69257		Q69257 bordetella
14	237	87.1	181	2	Q9ALQ1		Q9alq1 bordetella
15	235	86.4	912	2	Q8RSU0		Q8rsu0 bordetella
16	232.5	85.5	147	2	Q9K5H6		Q9k5h6 bordetella

```
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Pertactin (P.68) (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEL;
RX MEDLINE=20359389; PubMed=10899896;
RA Boursaux-Eude C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250088; CAB76442.1; -.
FT NON_TER 1
FT NON_TER 159
SQ SEQUENCE 159 AA; 15181 MW; 9F315BB478659E95 CRC64;

Query Match 100.0%; Score 272; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.5e-21;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 51
|||||
Db 53 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 103
|||||

RESULT 3
Q9K510 PRELIMINARY; PRT; 160 AA.
ID Q9K510;
AC Q9K510;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Pertactin (P.68) (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9.73H+;
RX MEDLINE=20359389; PubMed=10899896;
RA Boursaux-Eude C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250076; CAB76430.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 15280 MW; FE9618ECD81013EB CRC64;

Query Match 100.0%; Score 272; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 1.5e-21;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 51
|||||
Db 54 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 104
|||||

RESULT 4
Q9AHP1 PRELIMINARY; PRT; 179 AA.
ID Q9AHP1;
AC Q9AHP1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MBORD831;
RX MEDLINE=21117018; PubMed=11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella bronchiseptica Pertactin.";
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL; AF298589; AAK16690.1; -.
DR InterPro: IPR003992; pertactin.
DR PRINTS; PR01482; PERTACTIN.
FT NON_TER 1
FT NON_TER 179
SQ SEQUENCE 179 AA; 17270 MW; 325502FFB5483DF7 CRC64;

Query Match 100.0%; Score 272; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.7e-21;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 51
|||||
Db 79 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 129
|||||

RESULT 5
Q9KJY3 PRELIMINARY; PRT; 211 AA.
ID Q9KJY3;
AC Q9KJY3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Pertactin (Fragment).
GN OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B;
RA Keil D.J., Fenwick B.;
RT "Variability in the pertactin genes of Bordetella bronchiseptica isolates includes regions coding for the GGXXP and PQP families of amino acid repeats";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156769; AAF82393.1; -.
FT NON_TER 1
FT NON_TER 211
SQ SEQUENCE 211 AA; 20946 MW; E751FFA1510D99A4 CRC64;

Query Match 100.0%; Score 272; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 2e-21;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 51
|||||
Db 154 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDSTVDLAQ 204
|||||

RESULT 6
Q9KJY4 PRELIMINARY; PRT; 216 AA.
ID Q9KJY4;
AC Q9KJY4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Pertactin (Fragment).
GN OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
```


AC Q9S6M9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin precursor.
GN PRN4.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B705;
RX MEDLINE=99270984; PubMed=10338531;
RA Mooi F.R., He Q., Van Oirschot H., Mertsola J.;
RT "Variation in the Bordetella pertussis virulence factors pertussis toxin and pertactin in vaccine strains and clinical isolates in Finland.";
RL Infect. Immun. 67:3133-3134(1999).
DR EMBL; AJ011015; CAA09460.1; -;
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTNFAMILY.
KW Signal.
FT SIGNAL
FT CHAIN 35 905
SQ SEQUENCE 905 AA; 93071 MW; 7AE0F69C1B305E52 CRC64;

Query Match 88.6%; Score 241; DB 2; Length 905;
Best Local Similarity 88.2%; Pred. No. 1.8e-17;
Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 51
|||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 304
|||||

RESULT 11
Q9S6N0 PRELIMINARY; PRT; 905 AA.
AC Q9S6N0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin.
GN PRN8.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B1092;
RA van Oirschot H.F.L.M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-B1092;
RX MEDLINE=98114370; PubMed=9453625;
RA Mooi F.R., VanOirschot H., Heuvelman K., vanderHeide H., Gaastra W.,
RA Williams R.R.J.;
RT "Polymorphism in the Bordetella pertussis virulence factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution.";
RL Infect. Immun. 66:670-675(1998).
DR EMBL; AJ13245; CAB39891.1; -;
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.

DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTNFAMILY.
SQ SEQUENCE 905 AA; 93114 MW; 74B155EDB61059B4 CRC64;
Query Match 88.6%; Score 241; DB 2; Length 905;
Best Local Similarity 88.2%; Pred. No. 1.8e-17;
Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 51
|||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLAQ 304
|||||

RESULT 12
Q937U5 PRELIMINARY; PRT; 167 AA.
AC Q937U5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A67;
RA Weber C., Boursaux-Eude C., Nicole G.;
RT "Polymorphism of Bordetella pertussis isolates circulating the last ten years in France, a country using the same effective whole-cell vaccine since more than thirty years.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ318056; CAC67458.1; -;
FT NON_TER 1 167
FT NON_TER 167 167
SQ SEQUENCE 167 AA; 15826 MW; E136B4CF80955F0 CRC64;

Query Match 87.9%; Score 239; DB 2; Length 167;
Best Local Similarity 77.0%; Pred. No. 4.9e-18;
Matches 47; Conservative 3; Mismatches 1; Indels 10; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLA 50
|||||
Db 54 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWTGVDVSDTVDLA 113
|||||
QY 51 Q 51
Db 114 Q 114

RESULT 13
O69257 PRELIMINARY; PRT; 907 AA.
AC O69257;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin precursor.
GN PRN OR PRN6.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-18323;
RA Boursaux-Eude C., Guiso N.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-18323;

RA van Oirschot H.F.L.M., Mooi F.R.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CZ;
RX MEDLINE-99345256; PubMed-10418915;
RA Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;
RT "Intranasal murine model of Bordetella pertussis infection: II.
RT Sequence variation and protection induced by a tricomponent acellular
RT vaccine.";
RL Vaccine 17:2651-2660(1999).
DR EMBL: AJ006152; CAA06894.1; -
DR EMBL: AJ132095; CAA10566.1; -
DR EMBL: AJ006160; CAA06902.2; -
DR InterPro: IPR003992; Pertactin.
DR InterPro: IPR003991; pertactin_vir.
DR Pfam: PF03212; Pertactin_sup.
DR PRINTS: PR01482; PERTACTIN.
DR PRINTS: PR01484; PERTACTINFAMILY.
KW Signal.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 907 PERTACTIN.
SQ SEQUENCE 907 AA; 93298 MW; 3A7D05F4094420EA CRC64;

Query Match 87.3%; Score 237.5; DB 2; Length 907;
Best Local Similarity 82.3%; Pred. No. 4.3e-17;
Matches 46; Conservative 3; Mismatches 2; Indels 5; Gaps 1;
QY 1 QRATIRRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWDVSDSTVDLAQ 51
|||||
DB 254 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFGPGFVLDGWDVSGSSVELAQ 309
|||||

RESULT 14

Q9ALQ1
ID Q9ALQ1 PRELIMINARY; PRT; 181 AA.
AC Q9ALQ1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MBORD591;
RX MEDLINE-21117018; PubMed-11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella
RT bronchiseptica Pertactin."
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL: AY007266; AAG38442.1; -
DR InterPro: IPR003992; pertactin.
DR PRINTS: PR01482; PERTACTIN.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 17575 MW; 739181AE7294871A CRC64;

Query Match 87.1%; Score 237; DB 2; Length 181;
Best Local Similarity 86.3%; Pred. No. 8.7e-18;
Matches 44; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWDVSDSTVDLAQ 51
|||||
DB 83 QRATIRRGDAPAGGAVPGGAVPGGFGPGFVLDGWDVSGSSVELAQ 133
|||||

RESULT 15

Q8RSU0

ID Q8RSU0 PRELIMINARY; PRT; 912 AA.
AC Q8RSU0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pertactin.
GN PRN.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B442;
RA van Loo I.H., Mooi F.R.;
RT "Changes in the Bordetella pertussis population in the first 20 years
RT after the introduction of vaccination."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ430832; CAD23609.1; -
SQ SEQUENCE 912 AA; 93680 MW; 7F417BA66B732EF7 CRC64;

Query Match 86.4%; Score 235; DB 2; Length 912;
Best Local Similarity 75.4%; Pred. No. 7.9e-17;
Matches 46; Conservative 3; Mismatches 2; Indels 10; Gaps 1;
QY 1 QRATIRRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWDVSDSTVDLA 50
|||||
DB 254 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFGPGFVLDGWDVSGSSVELA 313
|||||
QY 51 Q 51
DB 314 Q 314

Search completed: May 7, 2003, 16:53:09
Job time: 29.0647 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 29.0971 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-9
Perfect score: 308
Sequence: 1 QRATIRGDAPAGGVPVGGAA.....PVLDGWGVDSGSTVELAQ 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	308	100.0	56	AAE16188	B. bronchiseptica
2	291	94.5	56	AAE16191	B. bronchiseptica
3	281	91.2	56	AAE16189	B. bronchiseptica
4	281	91.2	910	AAE16184	Bordetella pertussis
5	281	91.2	910	AAE17146	Bordetella pertussis
6	278.5	90.4	61	AAE16190	B. bronchiseptica
7	254.5	82.6	51	AAE16192	B. bronchiseptica
8	234	76.0	45	AAE17153	F647 monoclonal an
9	234	76.0	45	AAE17154	Pem3 monoclonal an
10	234	76.0	45	AAE17155	Pem4 monoclonal an

11	234	76.0	45	AAE17156	Pem68 monoclonal a
12	234	76.0	45	AAE17157	Pem70 monoclonal a
13	234	76.0	45	AAE17158	Pem71 monoclonal a
14	234	76.0	45	AAE17159	Pem72 monoclonal a
15	228.5	74.2	51	AAE16186	B. bronchiseptica
16	228.5	74.2	922	AAE25578	Bordetella paraper
17	228.5	74.2	922	AAE16185	Bordetella paraper
18	222.5	72.2	922	AAE14321	Pertactin antigen
19	212	58.8	46	AAE16187	B. bronchiseptica
20	212	58.8	911	AAE14320	Pertactin antigen
21	212	58.8	911	AAE26503	prn proteins. Bor
22	212	58.8	911	AAE16183	Bordetella bronchi
23	196	53.6	36	AAE17174	Maltose binding pr
24	191	52.0	59	AAE17149	Bordetella pertuss
25	190	51.7	36	AAE17172	Maltose binding pr
26	186	50.4	56	AAE17152	Bordetella pertuss
27	184.5	50.9	64	AAE17148	Bordetella pertuss
28	183.5	50.6	41	AAE17171	Maltose binding pr
29	181	50.4	59	AAE17147	Bordetella pertuss
30	180	50.4	36	AAE17141	Bordetella pertuss
31	180	50.4	36	AAE17170	Maltose binding pr
32	155.5	50.5	54	AAE17151	Bordetella pertuss
33	154.5	50.2	54	AAE17150	Bordetella pertuss
34	153.5	49.8	31	AAE17173	Maltose binding pr
35	105	34.1	20	AAE17144	Bordetella pertuss
36	104	33.8	445	ABB63068	Drosophila melanog
37	93	30.2	15	AAE17165	Monoclonal Ab deri
38	93	30.2	137	ABB69236	Drosophila melanog
39	91.5	29.7	334	AAE16181	Arabidopsis thalia
40	91.5	29.7	342	AB968951	Drosophila melanog
41	91.5	29.7	375	AAE16180	Arabidopsis thalia
42	91.5	29.7	441	AAE16179	Arabidopsis thalia
43	91.5	29.7	531	AAE140098	Spider silk protei
44	91.5	29.7	595	AAE14309	N.clavipes draglin
45	91.5	29.7	595	AAE53347	Nephila clavipes s

ALIGNMENTS

RESULT 1
AAE16188

ID AAE16188 standard; peptide; 56 AA.

AC AAE16188;

XX 26-MAR-2002 (first entry)

DT B. bronchiseptica strain I-3 pertactin outer membrane protein region I.

DE Pertactin; prn; outer membrane protein; vaccine; Bordetella infection;

XX therapy; antibiotic; antibacterial; region I.

OS Bordetella bronchiseptica.

XX WO200190143-A2.

XX 29-NOV-2001.

XX 23-MAY-2001; 2001WO-EP06457.

XX 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouze N, Boursaux-eude C;

XX WPI; 2002-097639/13.

PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

RESULT 4
AAE16184
ID AAE16184 standard; Protein; 910 AA.
AC AAE16184;
DT 26-MAR-2002 (first entry)
XX
DE Bordetella pertussis pertactin outer membrane protein, p.69.
XX
KW Pertactin; PRN: outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; p.69.
XX
OS Bordetella pertussis.
XX
FH Key Location/Qualifiers
FT Region 254..309
FT /note= "Pertactin region I"
FT Region 568..609
FT /note= "Pertactin region II"
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PE 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969P.
XX
PA (INSP) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
DR N-PSDB; AAD26441.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
PS Disclosure; Page 31; 47pp; English.
XX
CC The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. pertussis
CC pertactin outer membrane protein, p.69.
XX
SQ Sequence 910 AA;
Query Match 91.2%; Score 281; DB 23; Length 910;
Best Local Similarity 92.9%; Pred. No. 7.7e-21;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGPVLGDWYGVDSGSTVELAQ 56
|||||
DB 254 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGPVLGDWYGVDSGSTVELAQ 309
|||||
RESULT 5
AAE17146
ID AAE17146 standard; Protein; 910 AA.
XX
AC AAE17146;

XX
DT 18-APR-2002 (first entry)
XX
DE Bordetella pertussis pertactin (Prn1) protein.
XX
KW Pertactin; prn1; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response.
XX
OS Bordetella pertussis.
XX
FH Key Location/Qualifiers
FT Region 597..604
FT /note= "Conserved region"
XX
PN WO200200695-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-NL00493.
XX
PR 30-JUN-2000; 2000EP-0202309.
XX
PA (NEW-) NEDERLANDEN MIN WELZIJN.
XX
PI Mooi FR;
XX
DR WPI; 2002-139897/18.
XX
CC New polypeptides derived from Bordetella pertussis pertactin, useful as
CC a vaccine against infections caused by Bordetella strains, and other
CC infectious diseases of mammals, e.g. diphtheria, tetanus, or polio
XX
PS Claim 11; Page 35-38; 52pp; English.
XX
CC The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (prn1). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections
CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is B. pertussis prn1 protein.
XX
SQ Sequence 910 AA;
Query Match 91.2%; Score 281; DB 23; Length 910;
Best Local Similarity 92.9%; Pred. No. 7.7e-21;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGPVLGDWYGVDSGSTVELAQ 56
|||||
DB 254 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGPVLGDWYGVDSGSTVELAQ 309
|||||
RESULT 6
AAE16190
ID AAE16190 standard; peptide; 61 AA.
XX
AC AAE16190;
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain prn2 pertactin outer membrane protein region I.
XX
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region I.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.

```

XX PD 29-NOV-2001.
XX PF
XX PR 23-MAY-2001; 2001WO-EP06457.
XX PR 25-MAY-2000; 2000US-206969P.
XX PA (INSP ) INST PASTEUR.
XX PI Guiso-maclouf N, Boursaux-eude C;
XX DR WPI; 2002-097639/13.
XX PT Polypeptides containing polymorphisms of the repeated regions of
XX PT pertactin in Bordetella species, useful in immunogenic compositions for
XX PT treating infections caused by Bordetella and in diagnostic methods
XX PS Disclosure; Fig 1b; 47pp; English.
XX CC The present invention relates to Bordetella bronchiseptica pertactin
XX CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
XX CC vaccine. Pertactin antibody is useful for treating Bordetella infections
XX CC and used to detect Bordetella antigens in biological preparations or in
XX CC purifying corresponding proteins, glycoproteins or their mixtures when
XX CC used in affinity chromatographic columns. Pertactin is useful as antigens
XX CC to identify antibodies to Bordetella in materials such as human or other
XX CC animal tissue and human or other animal cells, as well as biological
XX CC fluids, such as human or other animal body fluids, including human sera,
XX CC and to determine the concentration of Ab in those materials. Thus the
XX CC antigens can be used for qualitative or quantitative determination of
XX CC Bordetella in a material. The present sequence is B. bronchiseptica
XX CC pertactin outer membrane protein region I.
XX SQ Sequence 61 AA;

Query Match 90.4%; Score 278.5; DB 23; Length 61;
Best Local Similarity 86.9%; Pred. No. 1.1e-21;
Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGVPGGAV-----PGGFDGPGGPGFVLDGWDVSGSTVELA 55
DB 1 QRATIRRGDAPAGGVPGGAVPGGPGGPGGPGGPGFVLDGWDVSGSTVELA 60

QY 56 Q 56
DB 61 Q 61

RESULT 7
AAE16192
ID AAE16192 standard; peptide; 51 AA.
AC AAE16192;
XX DT 26-MAR-2002 (first entry)
XX DE B. bronchiseptica strain prn4 pertactin outer membrane protein region I.
XX KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX KW therapy; antibiotic; antibacterial; region I.
XX OS Bordetella bronchiseptica.
XX PN WO200190143-A2.
XX PD 29-NOV-2001.
XX PF 23-MAY-2001; 2001WO-EP06457.
XX PR 25-MAY-2000; 2000US-206969P.
XX PA (INSP ) INST PASTEUR.
XX PT
XX PT

```

```

PI Guiso-maclouf N, Boursaux-eude C;
XX DR WPI; 2002-097639/13.
XX PT Polypeptides containing polymorphisms of the repeated regions of
XX PT pertactin in Bordetella species, useful in immunogenic compositions for
XX PT treating infections caused by Bordetella and in diagnostic methods
XX PS Disclosure; Fig 1b; 47pp; English.
XX CC The present invention relates to Bordetella bronchiseptica pertactin
XX CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
XX CC vaccine. Pertactin antibody is useful for treating Bordetella infections
XX CC and used to detect Bordetella antigens in biological preparations or in
XX CC purifying corresponding proteins, glycoproteins or their mixtures when
XX CC used in affinity chromatographic columns. Pertactin is useful as antigens
XX CC to identify antibodies to Bordetella in materials such as human or other
XX CC animal tissue and human or other animal cells, as well as biological
XX CC fluids, such as human or other animal body fluids, including human sera,
XX CC and to determine the concentration of Ab in those materials. Thus the
XX CC antigens can be used for qualitative or quantitative determination of
XX CC Bordetella in a material. The present sequence is B. bronchiseptica
XX CC pertactin outer membrane protein region I.
XX SQ Sequence 51 AA;

Query Match 82.6%; Score 254.5; DB 23; Length 51;
Best Local Similarity 87.5%; Pred. No. 2.7e-19;
Matches 49; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGVPGGAVPGGFDGPGGPGFVLDGWDVSGSTVELAQ.56
DB 1 QRATIRRGDAPAGGVPGGAV-----PGGFDGPGGPGFVLDGWDVSGSTVELAQ 51

RESULT 8
AAE17153
ID AAE17153 standard; Protein; 45 AA.
XX AC AAE17153;
XX DT 18-APR-2002 (first entry)
XX DE F647 monoclonal antibody specific for region 1 of pertactin.
XX KW Pertactin; prn1; antibacterial; immunostimulant; antimicrobial; vaccine;
XX KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
XX KW immune response; antibody.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Region 31..40
XX FT /note= "Epitope"
XX PN WO200200695-A2.
XX PD 03-JAN-2002.
XX PF 29-JUN-2001; 2001WO-NL00493.
XX PR 30-JUN-2000; 2000EP-0202309.
XX PA (NEWE-) NEDERLANDEN MIN WELZIJN.
XX PI MOOI FR;
XX DR WPI; 2002-139897/18.
XX PT New polypeptides derived from Bordetella pertussis pertactin, useful as
XX PT a vaccine against infections caused by Bordetella strains, and other
XX PT infectious diseases of mammals, e.g. diphtheria, tetanus, or polio
XX PT

```

PS Example 1; Fig 2a; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (Prnl). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections
CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is a monoclonal antibody specific to region 1 of pertactin
CC used for epitope mapping and vaccination.

XX Sequence 45 AA;

Query Match 76.0%; Score 234; DB 23; Length 45;

Best Local Similarity 93.3%; Pred. No. 3.1e-17;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATIRGDAPAGGVPGGVPGGFDGPGGFGPGVLDGWYGV 47

DB 1 ATIRGDAPAGGVPGGVPGGFDGPGGFGPGVLDGWYGV 45

RESULT 9

AAEL17154

ID AAE17154 standard; Protein; 45 AA.

XX AAE17154;

XX 18-APR-2002 (first entry)

DE Pem3 monoclonal antibody specific for region 1 of pertactin.

XX Pertactin; prnl; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response; antibody.

XX Unidentified.

XX Key Location/Qualifiers
FH Region 26..35

FT /note- "Epitope"

XX WO200200695-A2.

XX 03-JAN-2002.

XX 29-JUN-2001; 2001WO-NL00493.

XX 30-JUN-2000; 2000EP-0202309.

XX (NEW-) NEDERLANDEN MIN WELZIJN.

XX Mooi FR;

XX WPI; 2002-139897/18.

XX New polypeptides derived from Bordetella pertussis pertactin, useful as
PT a vaccine against infections caused by Bordetella strains, and other
PT infectious diseases of mammals, e.g. diphtheria, tetanus, or polio -

XX Example 1; Fig 2a; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (Prnl). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections

CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is a monoclonal antibody specific to region 1 of pertactin
CC used for epitope mapping and vaccination.

XX Sequence 45 AA;

Query Match 76.0%; Score 234; DB 23; Length 45;

Best Local Similarity 93.3%; Pred. No. 3.1e-17;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATIRGDAPAGGVPGGVPGGFDGPGGFGPGVLDGWYGV 47

DB 1 ATIRGDAPAGGVPGGVPGGFDGPGGFGPGVLDGWYGV 45

RESULT 10

AAEL17155

ID AAE17155 standard; Protein; 45 AA.

XX AAE17155;

XX 18-APR-2002 (first entry)

DE Pem4 monoclonal antibody specific for region 1 of pertactin.

XX Pertactin; prnl; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response; antibody.

XX Unidentified.

XX Key Location/Qualifiers
FH Region 26..35

FT /note- "Epitope"

XX WO200200695-A2.

XX 03-JAN-2002.

XX 29-JUN-2001; 2001WO-NL00493.

XX 30-JUN-2000; 2000EP-0202309.

XX (NEW-) NEDERLANDEN MIN WELZIJN.

XX Mooi FR;

XX WPI; 2002-139897/18.

XX New polypeptides derived from Bordetella pertussis pertactin, useful as
PT a vaccine against infections caused by Bordetella strains, and other
PT infectious diseases of mammals, e.g. diphtheria, tetanus, or polio -

XX Example 1; Fig 2a; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (Prnl). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections
CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is a monoclonal antibody specific to region 1 of pertactin
CC used for epitope mapping and vaccination.

XX Sequence 45 AA;

Query Match 76.0%; Score 234; DB 23; Length 45;

Best Local Similarity 93.3%; Pred. No. 3.1e-17;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 47
DB 1 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 45

RESULT 11

AAE17156
ID AAE17156 standard; Protein: 45 AA.

XX AC AAE17156;

XX DT 18-APR-2002 (first entry)

XX DE Pem68 monoclonal antibody specific for region 1 of pertactin.

XX KW Pertactin; prnl; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response; antibody.

XX OS Unidentified.

XX FH Key Location/Qualifiers
XX FT Region 6..15
XX FT /note= "Epitope"

XX PN WO200200695-A2.

XX PD 03-JAN-2002.

XX PF 29-JUN-2001; 2001WO-NL00493.

XX PR 30-JUN-2000; 2000EP-0202309.

XX PA (NEWE-) NEDERLANDEN MIN WELZIJN.

XX PI Mooi FR;

XX DR WPI; 2002-139897/18.

XX New polypeptides derived from Bordetella pertussis pertactin, useful as
PT a vaccine against infections caused by Bordetella strains, and other
PT infectious diseases of mammals, e.g. diphtheria, tetanus, or polio

XX PS Example 1; Fig 2a; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (Prnl). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections
CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is a monoclonal antibody specific to region 1 of pertactin
CC used for epitope mapping and vaccination.

XX SQ Sequence 45 AA;

Query Match 76.0%; Score 234; DB 23; Length 45;

Best Local Similarity 93.3%; Pred. No. 3.1e-17;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 47

DB 1 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 45

RESULT 12

AAE17157
ID AAE17157 standard; Protein: 45 AA.

XX AC AAE17157;

XX

XX DT 18-APR-2002 (first entry)

XX DE Pem70 monoclonal antibody specific for region 1 of pertactin.

XX KW Pertactin; prnl; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response; antibody.

XX OS Unidentified.

XX FH Key Location/Qualifiers
XX FT Region 1..5
XX FT /note= "Epitope"

XX PN WO200200695-A2.

XX PD 03-JAN-2002.

XX PF 29-JUN-2001; 2001WO-NL00493.

XX PR 30-JUN-2000; 2000EP-0202309.

XX PA (NEWE-) NEDERLANDEN MIN WELZIJN.

XX PI Mooi FR;

XX DR WPI; 2002-139897/18.

XX New polypeptides derived from Bordetella pertussis pertactin, useful as
PT a vaccine against infections caused by Bordetella strains, and other
PT infectious diseases of mammals, e.g. diphtheria, tetanus, or polio

XX PS Example 1; Fig 2a; 52pp; English.

XX The invention relates to polypeptides derived from Bordetella pertussis
CC pertactin (Prnl). The polypeptide is useful in the preparation of
CC vaccines against B. pertussis, B. parapertussis, B. bronchiseptica and
CC other infectious diseases of mammals including diphtheria, tetanus, polio
CC and infections caused by Haemophilus influenza b. The polypeptide is
CC especially useful for eliciting an immune response against Bordetella sp.
CC Antibodies against the polypeptide may be used for pharmaceutical and/or
CC diagnostic purposes, particularly for treating or preventing infections
CC caused by Bordetella pertussis or Bordetella parapertussis. The present
CC sequence is a monoclonal antibody specific to region 1 of pertactin
CC used for epitope mapping and vaccination.

XX SQ Sequence 45 AA;

Query Match 76.0%; Score 234; DB 23; Length 45;

Best Local Similarity 93.3%; Pred. No. 3.1e-17;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 47

DB 1 ATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGWTGVDV 45

RESULT 13

AAE17158

ID AAE17158 standard; Protein: 45 AA.

XX AC AAE17158;

XX DT 18-APR-2002 (first entry)

XX DE Pem71 monoclonal antibody specific for region 1 of pertactin.

XX KW Pertactin; prnl; antibacterial; immunostimulant; antimicrobial; vaccine;
KW diphtheria; tetanus; polio; Haemophilus influenza b infection; therapy;
KW immune response; antibody.

XX OS Unidentified.

CC The present invention relates to Bordetella bronchiseptica pertactin
 CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
 CC vaccine. Pertactin antibody is useful for treating Bordetella infections
 CC and used to detect Bordetella antigens in biological preparations or in
 CC purifying corresponding proteins, glycoproteins or their mixtures when
 CC used in affinity chromatographic columns. Pertactin is useful as antigens
 CC to identify antibodies to Bordetella in materials such as human or other
 CC animal tissue and human or other animal cells, as well as biological
 CC fluids, such as human or other animal body fluids, including human sera,
 CC and to determine the concentration of Ab in those materials. Thus the
 CC antigens can be used for qualitative or quantitative determination of
 CC Bordetella in a material. The present sequence is B. bronchiseptica
 CC pertactin outer membrane protein region I.
 XX
 SQ Sequence 51 AA;

Query Match 74.2%; Score 228.5; DB 23; Length 51;
 Best Local Similarity 80.4%; Pred. No. 1.3e-16;
 Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;
 QY 1 QRATIRRGDAPAGGVPGGVPGGFGPGFPGVLDGWYGVDSVSTVLAQ 56
 DB 1 QRATIRRGDAPAGGVPGGAV-----PGGAVPGGFGPLLDGWYGVDSVSTVLAQ 51

Search completed: May 7, 2003, 16:47:03
 Job time : 30.0971 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:41:51 ; Search time 9.25383 Seconds
(without alignments)
178.054 Million cell updates/sec

Title: US-09-855-754B-9

Perfect score: 308

Sequence: 1 QRATIRGDAPAGGVPVGA.....PVLGMYGVDSGSTVELAQ 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	88.0	910	4	US-08-460-269C-2
2	228.5	74.2	922	4	US-08-460-269C-6
3	212	68.8	911	4	US-08-460-269C-4
4	120.5	39.1	907	2	US-09-010-928B-4
5	116	37.7	745	2	US-09-010-928B-28
6	116	37.7	870	2	US-09-010-928B-2
7	91.5	29.7	595	1	US-08-425-069-4
8	91.5	29.7	595	2	US-08-317-844B-4
9	90.5	29.4	119	4	US-08-556-978B-60
10	90.5	29.4	119	4	US-09-247-806-9
11	90.5	29.4	714	4	US-08-556-978B-61
12	90.5	29.4	714	4	US-09-247-806-10
13	90	29.2	420	2	US-08-845-998-8
14	90	29.2	420	2	US-09-206-537-8
15	90	29.2	420	4	US-09-430-854-8
16	87.5	28.4	529	4	US-09-247-806-2
17	86.5	28.1	47	4	US-08-556-978B-18
18	86.5	28.1	493	4	US-08-556-978B-59
19	82	26.6	20	2	US-09-010-928B-9
20	79	25.6	20	2	US-09-010-928B-7
21	77	25.0	28	4	US-08-556-978B-58
22	77	25.0	32	1	US-08-425-069-34
23	77	25.0	32	1	US-08-209-747-45
24	77	25.0	32	1	US-08-458-298-45
25	77	25.0	32	2	US-08-317-844B-34
26	74	24.0	262	3	US-08-946-914-14
27	74	24.0	262	4	US-09-656-450-14

28	73.5	23.9	35	1	US-08-209-747-52	Sequence 52, Appl
29	73.5	23.9	35	1	US-08-458-298-52	Sequence 52, Appl
30	71.5	23.2	23	4	US-08-556-978B-46	Sequence 46, Appl
31	71	23.1	112	3	US-08-963-168C-16	Sequence 16, Appl
32	71	23.1	132	3	US-08-963-168C-9	Sequence 9, Appl
33	70	22.7	264	1	US-08-562-311-4	Sequence 4, Appl
34	70	22.7	315	1	US-08-257-073-7	Sequence 7, Appl
35	70	22.7	646	1	US-08-441-139-14	Sequence 14, Appl
36	70	22.7	890	4	US-09-513-783A-174	Sequence 174, App
37	69.5	22.6	43	3	US-08-542-051-31	Sequence 31, Appl
38	69.5	22.6	43	3	US-08-542-051-32	Sequence 32, Appl
39	69.5	22.6	43	3	US-08-542-051-33	Sequence 33, Appl
40	69.5	22.6	48	3	US-08-542-051-45	Sequence 45, Appl
41	69.5	22.6	160	3	US-08-542-051-18	Sequence 18, Appl
42	69.5	22.6	251	1	US-08-397-633A-74	Sequence 74, Appl
43	69.5	22.6	466	3	US-08-526-136-13	Sequence 13, Appl
44	69.5	22.6	479	1	US-08-397-633A-78	Sequence 78, Appl
45	69.5	22.6	486	1	US-08-397-633A-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-08-460-269C-2
; Sequence 2, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 910 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2

Query Match 88.0%; Score 271; DB 4; Length 910;
Best Local Similarity 91.1%; Pred. No. 5.6e-22;
Matches 51; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGVPVGA...PVLGMYGVDSGSTVELAQ 56
DB 254 QRATIRGDALAGGAVPGGVPVGA...PVLGMYGVDSGSTVELAQ 309

```

RESULT 2
US-08-460-269C-6
; Sequence 6, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6

Query Match 74.2%; Score 228.5; DB 4; Length 922;
Best Local Similarity 80.4%; Pred. No. 2.5e-17;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGVGVDSVSTVELAQ 56
|||||
Db 254 QRATIRGDAPAGGVPGGVPGGFGPGLDGVGVDSVSTVDLAQ 304
|||||-----PGGAVPGGFGPGLDGVGVDSVSTVDLAQ 304

RESULT 3
US-08-460-269C-4
; Sequence 4, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08460269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-460-269C-4

Query Match 68.8%; Score 212; DB 4; Length 911;
Best Local Similarity 75.0%; Pred. No. 1.6e-15;
Matches 42; Conservative 2; Mismatches 2; Indels 10; Gaps 1;

QY 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGVGVDSVSTVELAQ 56
|||||
Db 254 QRATIRGDAPAGGVPGGVPGGFGPGLDGVGVDSVSTVDLAQ 299
|||||-----PGGFGPGLDGVGVDSVSTVDLAQ 299

RESULT 4
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-010-928B-4

Query Match 39.1%; Score 120.5; DB 2; Length 907;
Best Local Similarity 48.3%; Pred. No. 1.7e-05;
Matches 28; Conservative 1; Mismatches 18; Indels 11; Gaps 1;

QY 4 TTRRGDAP-----AGGVPGGVPGGFGPGLDGVGVDSVSTVELAQ 50
|||||
Db 668 TIDGADGPTTISELTISAGGSGPGGAGPGGVPGGVPGGVPGGVPGGVPGGV 725
|||||-----AGGVPGGVPGGFGPGLDGVGVDSVSTVELAQ 50

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-2
Query Match          37.7%; Score 91.6; DB 2; Length 870;
Best Local Similarity 56.5%; Pred. No. 5e-05;
Matches 26; Conservative 3; Mismatches 14; Indels 3; Gaps 1

QY      8 GDAPAGGV-PGGAVPGGDPGGFGPGF--VLDGWYGVDVSGS 50
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     702 GGAGSGVGPGGSGPGGYPGGAGPGGYGGSGGPGGSGS 747

RESULT 7
US-08-425-069-4
; Sequence 4, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22045
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-069-4
Query Match          29.7%; Score 91.5; DB 1; Length 595;
Best Local Similarity 41.5%; Pred. No. 0.016;
Matches 22; Conservative 3; Mismatches 25; Indels 3; Gaps 1

QY      3 ATTRGDPAAGCGVPCGAVPGGFDPGGFGPGGVPVLDGWYGVDSGSTVELA 55
       | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
Db    119 ASAESQQQPPGYPGQGQGGPGGPGGPGGPGGP---QQQPGSPGPGSAAAAA 168

```

RESULT 9
US-08-556-978B-60
; Sequence 60, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH

```

Query Match          29.4%; Score 90.5; DB 4; Length 119;
Best Local Similarity 43.8%; Pred. No. 0.0037;
Matches 21; Conservative 2; Mismatches 22; Indels 3; Gaps 1;

QY      8  GDAPAGGVPGGAVPGGFDPCGGFGPGGPGVLDGWWGVVDVSGSTVELA 55
      | | | | | | | | | | | | | | | | | | | | |
Db      6  GSAAAAAGPQCGPGGYPGQCGQPGGYP--GQCGPSPGCSAAAA 50

RESULT 11
US-08-556-978B-61
; Sequence 61, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:

```

```

Query Match          29.4%; Score 90.5; DB 4; Length 7
Best Local Similarity 43.8%; Pred. No. 0.025;
Matches 21; Conservative 2; Mismatches 22; Indels

QY      8  GDAPAGGVPGGVPGGFGDGGCGPGGFGPVLGDWGYGVDSGSGVELA 55
Db      601  GSAAAAAGPQQGPGGFGPQQGPGGYGP---GQQGPGPGCSAAAA 645

RESULT 13
US-08-845-998-8
; Sequence 8, Application US/08845998
; Patent No. 5873692
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Hoon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,998

```

Query Match 29.4%; Score 90.5; DB 4; Length 714;
Best Local Similarity 43.8%; Pred. No. 0.025;
Matches 21; Conservative 2; Mismatches 22; Indels 3; Caps 1;

```

; . . . LENGTH: 420 amino acids
; . . . TYPE: amino acid
; . . . TOPOLOGY: linear
; . . . MOLECULE TYPE: protein
US-08-845-998-8

Query Match      29.2%; Score 90; DB 2; Length 420;
Best Local Similarity 60.0%; Pred. No. 0.016;
Matches 18; Conservative 0; Mismatches 12; Indels

QY      8  GDAPAGGVPGGAVPGGFDPGGFGPGGGF 37
      | | | | | | | | | | | | | |
Db      150  GGGPGGGGCGPGGGGPGGGGGGPGGGGGP 179

RESULT 14
US-09-206-537-8
; Sequence 8, Application US/09206537
; Patent No. 6130052
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES

```

```

; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/206,537
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-206-537-8

Query Match 29.2%; Score 90; DB 4; Length 420;
Best Local Similarity 60.0%; Pred. No. 0.016;
Matches 18; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GDAPAGGVPGGAVPGGFGPGGFGPG 37
Db 150 GGGPGGGGPGGGGPGGGGPGGGGPG 179

RESULT 15
US-09-430-854-8
; Sequence 8, Application US/09430854
; Patent No. 6271019
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,854
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-430-854-8

Query Match 29.2%; Score 90; DB 4; Length 420;
Best Local Similarity 60.0%; Pred. No. 0.016;
Matches 18; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GDAPAGGVPGGAVPGGFGPGGFGPG 37
Db 150 GGGPGGGGPGGGGPGGGGPGGGGPG 179

Search completed: May 7, 2003, 16:56:58
Job time : 10.3447 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:53:34 ; Search time 16.7905 Seconds
(without alignments)
306.927 Million cell updates/sec

Title: US-09-855-754B-9
Perfect score: 308
Sequence: 1 QRATIRGDAPAGGGVPGCA.....PVLGDWGYVDVSGSTVELAQ 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	308	100.0	56	9	US-09-855-754-9
2	291	94.5	56	9	US-09-855-754-12
3	281	91.2	56	9	US-09-855-754-10
4	278.5	90.4	61	9	US-09-855-754-11
5	274	89.0	91.0	9	US-09-855-754-5
6	254.5	82.6	51	9	US-09-855-754-13
7	228.5	74.2	51	9	US-09-855-754-7
8	228.5	74.2	92.2	9	US-09-855-754-6
9	212	68.8	46	9	US-09-855-754-8
10	212	68.8	91.1	9	US-09-855-754-4
11	90.5	29.4	119	10	US-09-861-597-9
12	90.5	29.4	74.0	10	US-09-861-597-10
13	88	28.6	450	10	US-09-812-382-6
14	87.5	28.4	529	10	US-09-861-597-2
15	82	26.6	618	10	US-09-925-300-1381
16	79.5	25.8	47	10	US-09-040-518-3
17	78	25.3	50	10	US-09-812-382-3
18	78	25.3	382	9	US-09-738-626-6012
19	77.5	25.2	48	10	US-09-837-969A-9

ALIGNMENTS

RESULT 1

US-09-855-754-9

; Sequence 9, Application US/09855754

; Publication No. US2002019237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX-UEDE, CAROLINE

; APPLICANT: GUISSO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,

; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754

; CURRENT FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/206,969

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 56

; TYPE: PRT

; ORGANISM: Bordetella bronchiseptica

; US-09-855-754-9

Query Match 100.0%; Score 308; DB 9; Length 56;

Best Local Similarity 100.0%; Pred. No. 5.4e-24;

Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QRATIRGDAPAGGGVPGGVPDGGFGGPGVLDGWGYVDVSGSTVELAQ 56

|||||

Db 1 QRATIRGDAPAGGGVPGGVPDGGFGGPGVLDGWGYVDVSGSTVELAQ 56

|||||

RESULT 2

US-09-855-754-12

; Sequence 12, Application US/09855754

; Publication No. US2002019237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX-UEDE, CAROLINE

; APPLICANT: GUISSO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000
 ; CURRENT APPLICATION NUMBER: US/09/855,754
 ; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/206,969
 ; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 12
 ; LENGTH: 56
 ; TYPE: PRT
 ; ORGANISM: Bordetella bronchiseptica
 US-09-855-754-12

Query Match 94.5%; Score 291; DB 9; Length 56;
 Best Local Similarity 94.6%; Pred. No. 2.6e-22;
 Matches 53; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 56
 Db 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 56

RESULT 3
 US-09-855-754-10
 ; Sequence 10, Application US/09855754
 ; Publication No. US20020192237A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOURSAX-UEDE, CAROLINE
 ; APPLICANT: GUISSO-MACLOUF, NICOLE
 ; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
 ; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000
 ; CURRENT APPLICATION NUMBER: US/09/855,754
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/206,969
 ; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 56
 ; TYPE: PRT
 ; ORGANISM: Bordetella bronchiseptica
 US-09-855-754-10

Query Match 91.2%; Score 281; DB 9; Length 56;
 Best Local Similarity 92.9%; Pred. No. 2.5e-21;
 Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 56
 Db 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 56

RESULT 4
 US-09-855-754-11
 ; Sequence 11, Application US/09855754
 ; Publication No. US20020192237A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOURSAX-UEDE, CAROLINE
 ; APPLICANT: GUISSO-MACLOUF, NICOLE
 ; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
 ; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/206,969
 ; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 61
 ; TYPE: PRT
 ; ORGANISM: Bordetella bronchiseptica
 US-09-855-754-11

Query Match 90.4%; Score 278.5; DB 9; Length 61;
 Best Local Similarity 86.9%; Pred. No. 4.8e-21;
 Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;
 Qy 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELA 55
 Db 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELA 60
 Qy 56 Q 56
 Db 61 Q 61

RESULT 5
 US-09-855-754-5
 ; Sequence 5, Application US/09855754
 ; Publication No. US20020192237A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOURSAX-UEDE, CAROLINE
 ; APPLICANT: GUISSO-MACLOUF, NICOLE
 ; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
 ; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000
 ; CURRENT APPLICATION NUMBER: US/09/855,754
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/206,969
 ; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 910
 ; TYPE: PRT
 ; ORGANISM: Bordetella pertussis
 US-09-855-754-5

Query Match 89.0%; Score 274; DB 9; Length 910;
 Best Local Similarity 91.1%; Pred. No. 1.7e-19;
 Matches 51; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 56
 Db 254 QRATIRGDAPAGGAVPGGPGGFGPGLDGVGVDSGSTVELAQ 309

RESULT 6
 US-09-855-754-13
 ; Sequence 13, Application US/09855754
 ; Publication No. US20020192237A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOURSAX-UEDE, CAROLINE
 ; APPLICANT: GUISSO-MACLOUF, NICOLE
 ; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
 ; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000
 ; CURRENT APPLICATION NUMBER: US/09/855,754

;; CURRENT FILING DATE: 2001-09-10
;; PRIOR APPLICATION NUMBER: 60/206,969
;; PRIOR FILING DATE: 2000-05-25
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-13

Query Match 82.6%; Score 254.5; DB 9; Length 51;
Best Local Similarity 87.5%; Pred. No. 9.5e-19;
Matches 49; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

Qy 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGGVGVGVDSGSTVELAQ 56
|||||
Db 1 QRATIRGDAPAGGAVPGGAV-----PGGFGPGLDGGVGVDSGSTVELAQ 51

RESULT 7

US-09-855-754-7
; Sequence 7, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-7

Query Match 74.2%; Score 228.5; DB 9; Length 51;
Best Local Similarity 80.4%; Pred. No. 3.5e-16;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

Qy 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGGVGVGVDSGSTVELAQ 56
|||||
Db 1 QRATIRGDAPAGGAVPGGAV-----PGGFGPGLDGGVGVDSGSTVELAQ 51

RESULT 8

US-09-855-754-6
; Sequence 6, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-855-754-6

Query Match 74.2%; Score 228.5; DB 9; Length 922;
Best Local Similarity 80.4%; Pred. No. 5.3e-15;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

Qy 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGGVGVGVDSGSTVELAQ 56
|||||
Db 254 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFGPGLDGGVGVDSGSTVELAQ 304

RESULT 9

US-09-855-754-8
; Sequence 8, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-8

Query Match 68.8%; Score 212; DB 9; Length 46;
Best Local Similarity 75.0%; Pred. No. 1.4e-14;
Matches 42; Conservative 2; Mismatches 2; Indels 10; Gaps 1;

Qy 1 QRATIRGDAPAGGVPGGVPGGFGPGLDGGVGVGVDSGSTVELAQ 56
|||||
Db 1 QRATIRGDAPAGGAVPGGAV-----PGGFGPGLDGGVGVDSGSTVELAQ 46

RESULT 10

US-09-855-754-4
; Sequence 4, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-4

Query Match	29.4%	Score 90.5	DB 10	Length 714
Best Local Similarity	43.8%	Pred. No. 0.18		
Matches 21	Conservative 2	Mismatches 22	Indels 3	Gaps 1
8	GDAPAGGVP	GGAVPGG	FGPDPGG	FGFVLGWTGVDSGTVELA 55
601	GSAAAAAG	FGQGGY	FGQGGY	GGYGP---GGQGPSGSAANA 645

```

US-09-812-382-6
; Sequence 6, Application US/09812382
; Patent No. US20010034050A1
; GENERAL INFORMATION:
; APPLICANT: Chilkoti, Ashutosh
; TITLE OF INVENTION: Fusion Peptides Isolatable by Phase Transition
; FILE REFERENCE: 4176-101
; CURRENT APPLICATION NUMBER: US/09/812,382
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/190,659
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 450
; TYPE: PRT
; ORGANISM: ELP[V5A2G3-90]
US-09-812-382-6

```

RESULT 14
US-09-861-597-2
; Sequence 2, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:

```

; APPLICANT:  ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION:  COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION:  ANALOG
; FILE REFERENCE:  6388-0365-0
; CURRENT APPLICATION NUMBER:  US/09/861,597
; CURRENT FILING DATE:  2001-05-22
; PRIOR APPLICATION NUMBER:  09/247,806
; PRIOR FILING DATE:  1999-02-11
; PRIOR APPLICATION NUMBER:  FR 98/01614
; PRIOR FILING DATE:  1998-02-11
; NUMBER OF SEQ ID NOS:  14
; SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH:  529
; TYPE:  PRT
; ORGANISM:  Nephila clavipes
US-09-861-597-2

Query Match      28.4%; Score 87.5; DB 10; Length 529;
Best Local Similarity 43.3%; Pred. No. 0.26;
Matches 20; Conservative 3; Mismatches 20; Indels 3; Gaps 1;

QY      10  APAGGGVPGGAVPGGPPGGFGPGGVLDGWTGVDVSGSTVELA 55
          | | | | | | | | | | | | | | | | | | | | | |
          | | | | | | | | | | | | | | | | | | | | | |
          | | | | | | | | | | | | | | | | | | | | | |

db      206  AAAAAAGPGGGGPGGPGGPGGCGYCP---GQOGLSGPGSAAAAA 248
          | | | | | | | | | | | | | | | | | | | | | |
          | | | | | | | | | | | | | | | | | | | | | |
          | | | | | | | | | | | | | | | | | | | | | |

```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	271	88.0	910	2	A32560	outer membrane pro	
2	228.5	74.2	922	2	S15204	pertactin - Bordet	
3	212	68.8	911	2	A47675	58k outer membrane	
4	106	34.4	783	2	T35389	probable serine-th	
5	95.5	31.0	76	2	T02622	glycine-rich cell	
6	91.5	29.7	627	2	A44112	spidiroin 2, dragli	
7	91	29.5	860	1	EWMS	elastin precursor	
8	90	29.2	200	2	B86181	elastin precursor	
9	90	29.2	420	2	I59234	hypothetical prote	
10	89.5	29.1	646	2	T54677	octamer binding tr	
11	85	27.6	273	2	T40618	ATP-dependent RNA	
12	84	27.3	259	2	T5126	probable cell wall	
13	82.5	26.8	641	2	PC7036	hypothetical prote	
14	82.5	26.8	645	2	S41372	heat shock protein	
15	82	26.6	133	2	H84217	dnak-type molecule	
16	82	26.6	561	2	A31994	hypothetical prote	
17	82	26.6	593	1	KRH00	keratin 10, type I	
18	82	26.6	2918	2	A54105	keratin 10, type I	
19	81.5	26.5	422	2	T24865	fibrillin-2 precu	
20	81	26.3	686	2	A49242	hypothetical prote	
21	80.5	26.1	224	2	T51742	dnak-type molecule	
22	80.5	26.1	603	2	T45671	RNA helicase RH11	
23	80.5	26.1	864	1	EAT3	ATP-dependent RNA	
24	80.5	26.1	1300	2	T03166	elastin precursor	
25	80	26.0	656	2	S51712	probable immediate	
26	80	26.0	784	2	A26601	dnak-type molecule	
27	79.5	25.8	235	2	G60126	elastin precursor	
28	78.5	25.5	128	2	S25103	leukophysin - huma	
29	78	25.3	198	2	A54507	MFS18 protein - ma	
30	78	25.3	198	2	A54507	dnak-type molecule	

Qy

R.Beck, C.; Wamsley, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of *C. elegans* cosmid T20B6.
A:Reference number: Z18297
A:Accession: T15126
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-259 <BEC>
A:Cross-references: EMBL:AF000193; NID:g1946986; PID:g1946988; PIDN:AAB52890.1; GSPDB:GN
A:Experimental source: strain Bristol N2; clone T20B6
C:Genetics:
A:Gene: CESP-T20B6.3
A:Map position: 3
A:Introns: 9/2; 231/1
C:Superfamily: Phaseolus glycinine-rich protein 1.0

Query Match 27.3%; Score 84; DB 2; Length 259;
Best Local Similarity 44.7%; Pred. No. 0.35;
Matches 21; Conservative 4; Mismatches 14; Indels 8; Gaps 3;

QY 8 GDAPAGGGVPGGAVPGGFGGFGP-----GGFGPVLGWDYGVDSG 49
Db 202 GCGYGGGGGGGGGGGGGCGGPGGCGYGGGGGGGPG--GGYGGGGGGG 245

RESULT 13
PC7036
heat shock protein 70 - *Rhizopus nigricans* (fragment)
C:Species: *Rhizopus nigricans*
C:Date: 03-Nov-2000 #sequence_revision 03-Nov-2000 #text_change 17-Nov-2000
C:Accession: PC7036
R:Cernila, B.; Cresnar, B.; Breskvar, K.
Biochem. Biophys. Res. Commun. 265, 494-498, 1999
A:Title: Induction of Hsp70 in the fungus *Rhizopus nigricans*.
A:Reference number: JC7132; MUID:20025372; PMID:10558896
A:Accession: PC7036
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-641 <CER>
A:Cross-references: GB:AF188289
C:Superfamily: heat shock protein 70

Query Match 26.8%; Score 82.5; DB 2; Length 641;
Best Local Similarity 68.2%; Pred. No. 1.2;
Matches 15; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 8 GDAPAGGGVPGGAVPGGFGGFGP 29
Db 609 GCMPPGGGPPGGGAPGPF-PGG 629

RESULT 14
S41372
dnaK-type molecular chaperone HSA1 - yeast (*Pichia angusta*)
N:Alternate names: heat shock protein hsp72
C:Species: *Pichia angusta*
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Apr-2000
C:Accession: S41372; S71640; S71636
R:Diesel, A.A.; Roggenkamp, R.R.
submitted to the EMBL Data Library, January 1994
A:Description: HSP70 genes of the yeast *Hansenula polymorpha*.
A:Reference number: S41372
A:Accession: S41372
A:Molecule type: DNA
A:Residues: 1-645 <DIE>
A:Cross-references: EMBL:Z29379; NID:g443914; PIDN:CAAB2570.1; PID:g443915
R:Titorenko, V.I.; Evers, M.E.; Diesel, A.; Samyn, B.; van Beeumen, J.; Roggenkamp, R.;
Yeast 12, 849-857, 1996
A:Title: Identification and characterization of cytosolic *Hansenula polymorpha* proteins
A:Reference number: S71636; MUID:96437974; PMID:8840502
A:Accession: S71640
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-645 <TIT>
A:Cross-references: EMBL:Z29379; NID:g443914; PIDN:CAAB2570.1; PID:g443915
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1994
A:Accession: S71636
A:Molecule type: protein
A:Residues: 4-10; 137-143; 158-164; 221-227; 329-335; 362-368; 388-394; 425-431; 573-579; 609-
C:Genetics:
A:Gene: HSA1; HSP72
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein co
C:Superfamily: heat shock protein 70
C:Keywords: ATP; cytosol; heat shock; molecular chaperone; stress-induced protein

Query Match 26.8%; Score 82.5; DB 2; Length 645;
Best Local Similarity 48.9%; Pred. No. 1.2;
Matches 22; Conservative 1; Mismatches 11; Indels 11; Gaps 3;

QY 9 DAPAGGGVPGGAVPGGFGGFGPVLGWDYGVDSGSTVE 53
Db 609 DLYAAGGVPGGAAAGGFP-----GAGGAAP-----GAD-QGPSVE 642

RESULT 15
H84217
hypothetical protein Vng0597h [imported] - *Halobacterium* sp. NRC-1
C:Species: *Halobacterium* sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84217
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of *Halobacterium* species NRC-1
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: H84217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <STO>
A:Cross-references: GB:AE004437; NID:gi0580192; PIDN:AAG19108.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0597H

Query Match 26.6%; Score 82; DB 2; Length 133;
Best Local Similarity 42.6%; Pred. No. 0.29;
Matches 26; Conservative 3; Mismatches 22; Indels 10; Gaps 3;

QY 6 RRGDAPAGGGVP---GGAVPGGFGP-----GGF---CPGCGFPGVLDGWDYGVDSGSTVELA 55
Db 47 RMADAWGGGGGPPGMGGGGGGGPPGMMGGMMGGGGGPPGMMGGGGGGGGTSETA 106

QY 56 Q 56
Db 107 E 107

Search completed: May 7, 2003, 16:55:20
Job time: 11.494 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.43782 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-9
Perfect score: 308
Sequence: 1 QRATIRGDAPAGGGVPGCA.....PVLGDWGVDSGSTVELAQ 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	281	91.2	910	1 PERT_BORPE	P14283 bordetella
2	228.5	74.2	922	1 PERT_BORPA	P24328 bordetella
3	212	68.8	911	1 PERT_BORBR	Q03035 bordetella
4	91.5	29.7	627	1 SPD2_NEPCL	P46804 nephila cla
5	91	29.5	860	1 ELS_MOUSE	P54320 mus musculus
6	90	29.2	423	1 BR3A_HUMAN	Q01851 homo sapien
7	85.5	27.8	421	1 BR3A_MOUSE	P17208 mus musculus
8	82.5	26.8	644	1 HS71_PICAN	P53421 pichia angu
9	82	26.6	593	1 K1CJ_HUMAN	P13645 homo sapien
10	82	26.6	2911	1 FBN2_HUMAN	P35556 homo sapien
11	81.5	26.5	1380	1 DDH9_MOUSE	O70133 mus musculus
12	81	26.3	686	1 HS70_PLACB	Q05746 plasmodium
13	80.5	26.1	649	1 HS70_BLAEM	P48720 blastocycladi
14	80.5	26.1	864	1 ELS_RAT	Q99372 rattus norv
15	80	26.0	655	1 HS71_CANAL	P41797 candida alb
16	80	26.0	750	1 ELS_CHICK	P07916 gallus gall
17	79.5	25.8	1270	1 DDH9_HUMAN	O08211 homo sapien
18	79	25.6	643	1 K2C1_HUMAN	P04264 homo sapien
19	78.5	25.5	128	1 HS70_MAIZE	P32439 zea mays (m
20	78	25.3	198	1 HS70_SCHJA	P12795 schistosoma
21	78	25.3	252	1 GRP1_PHAVU	P10495 phaseolus v
22	78	25.3	648	1 HS71_PUCGR	Q01877 puccinia gr
23	76.5	24.8	295	1 LEG3_CANFA	P38486 canis fami
24	76.5	24.8	627	1 HYDL_SFRCO	P42534 streptomyc
25	76.5	24.8	650	1 HS7C_BOVIN	P19120 bos taurus
26	76	24.7	384	1 GRP1_PETHY	P09789 petunia hyb
27	75.5	24.5	342	1 ROAL_SCHAM	P21522 schistocerc
28	75.5	24.5	638	1 K22O_HUMAN	Q01546 homo sapien
29	75	24.4	747	1 ELS_HOVIN	P04985 bos taurus
30	74.5	24.2	338	1 GRP_ARATH	P27483 arabidopsis
31	74	24.0	261	1 LEG3_RAT	P08699 rattus norv
32	74	24.0	526	1 K1CJ_BOVIN	P06394 bos taurus
33	73.5	23.9	645	1 K22E_HUMAN	P35908 homo sapien

ALIGNMENTS

RESULT 1

ID	PERT_BORPE	STANDARD;	PRT;	910 AA.
AC	P14283;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.69) (P.93).			
GN	PRN OR OMP69A.			
OS	Bordetella pertussis.			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=520;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=CN2952;			
RX	MEDLINE=89264462; PubMed=2542937;			
RA	Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,			
RA	Novotny P., Morrissey P., Fairweather N.F.;			
RT	"Molecular cloning and characterization of protective outer membrane			
RT	protein P.69 from Bordetella pertussis.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).			
RN	[2]			
RP	REVISIONS TO 264 AND 332.			
RX	MEDLINE=92407514; PubMed=1527510;			
RA	Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;			
RT	"Cloning, nucleotide sequence and heterologous expression of the			
RT	protective outer-membrane protein P.68 pertactin from Bordetella			
RT	bronchiseptica.";			
RL	J. Gen. Microbiol. 138:1697-1705(1992).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).			
RX	MEDLINE=96196517; PubMed=8609998;			
RA	Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;			
RT	"Structure of Bordetella pertussis virulence factor P.69 pertactin.";			
RL	Nature 381:90-92(1996).			
CC	-1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS			
CC	MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN			
CC	BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SUBCELLULAR LOCATION: Outer membrane.			
CC	-1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)			
CC	CONCENTRATION.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J04560; AAA22980.1; ALT_SEQ.			
DR	PIR; A32560; A32560.			
DR	InterPro; IPR004899; Pertactin_sup.			

P37705 daucus caro
P12035 homo sapien
Q92125 xenopus lae
P87047 paracoccidi
Q61555 mus musculu
P47953 cricetus
P19013 homo sapien
Q92804 homo sapien
Q90473 brachydanio
P10569 acanthamoeb
P10591 saccharomyc
O53553 mycobacteri

DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMLY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 910 P.93.
FT CHAIN 35 711 PERTACTIN (P.69).
FT PROPEP 712 910 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
ADHESION TO VARIOUS EUKARYOTIC CELL
LINES).
FT DOMAIN 266 290 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3.
FT REPEAT 281 285 4 (APPROXIMATE).
FT REPEAT 286 290 5 (APPROXIMATE).
FT DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.
SQ SEQUENCE 910 AA; 93452 MW; AL69871E20A2E7DB CRC64;
Query Match 91.2%; Score 281; DB 1; Length 910;
Best Local Similarity 92.9%; Pred. No. 1.5e-18;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPGFGLDGMVGVDSSTVELAQ 56
|||||
DB 254 QRATIRGDAPAGGAVPGGAVPGGFGPGFGLDGMVGVDSSTVELAQ 309
|||||
RESULT 2
PERT_BORPA STANDARD; PRT; 922 AA.
ID PERT_BORPA
AC P24378;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.70) (P.95).
GN PRN.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CN2591;
RX MEDLINE=91251771; PubMed=2041476;
RA Li L.J., Dougan G., Novotny P., Charles I.G.;
RT "P.70 pertactin, an outer-membrane protein from Bordetella
parapertussis: cloning, nucleotide sequence and surface expression in
Escherichia coli.";
RT Mol. Microbiol. 5:409-417(1991).
RL
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CONCENTRATIONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.
CC
CC EMBL; X54547; CAA38419.1;
CC EMBL; A26124; CAA01786.1;
CC EMBL; A19182; CAA01454.1;

PIR; S15204; S15204.
PIR; S14659; S14659.
DR InterPro; IPR004899; Pertactin_sup.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMLY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 922 P.95.
FT CHAIN 35 647 PERTACTIN (P.70).
FT PROPEP 648 922 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
ADHESION TO VARIOUS EUKARYOTIC CELL
LINES).
FT DOMAIN 266 290 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3.
FT REPEAT 281 285 4 (APPROXIMATE).
FT DOMAIN 575 603 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
SQ SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
Query Match 74.2%; Score 228.5; DB 1; Length 922;
Best Local Similarity 80.4%; Pred. No. 8.6e-14;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;
QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPGFGLDGMVGVDSSTVELAQ 56
|||||
DB 254 QRATIRGDAPAGGAVPGGAVPGGFGPGFGLDGMVGVDSSTVELAQ 304
|||||
RESULT 3
PERT_BORBR STANDARD; PRT; 911 AA.
ID PERT_BORBR
AC Q03035;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.68) (P.94).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.
RC STRAIN=CN7531;
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
protective outer-membrane protein P.68 pertactin from Bordetella
bronchiseptica.";
RT J. Gen. Microbiol. 138:1697-1705(1992).
RL
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CONCENTRATIONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.
CC
CC EMBL; X54815; CAA38584.1;


```
Query Match      29.5%; Score 91; DB 1; Length 860;
Best Local Similarity 40.4%; Pred. No. 0.23;
Matches 23; Conservative 6; Mismatches 20; Indels 8; Gaps 2;

QY 2 RATIRRGDAPAGGVPG-----GAVPGGFD----PGGFGGPGGPGVLDGWYGVDSVSGS 50
DB 534 KAAAKAGLPGVGGVPGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 590

RESULT 6
BR3A_HUMAN
ID AC Q01851; Q15318; STANDARD; PRT; 423 AA.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Brain-specific homeobox/POU domain protein 3A (BRN-3A) (OCT-T1)
DE (Homeobox/POU domain protein RDC-1).
GN POU4F1 OR BRN3A OR RDC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=95348789; PubMed=7623109;
RA Xiang M., Zhou L.-J., Macke J.P., Yoshioka T., Hendry S.H., Eddy R.L.,
RA Shows T.B., Nathans J.;
RT "The Brn-3 family of POU-domain factors: primary structure, binding
RT specificity, and expression in subsets of retinal ganglion cells and
RT somatosensory neurons.";
RL J. Neurosci. 15:4762-4785(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94052142; PubMed=8234287;
RA Bhargava A.K., Li Z., Weissman S.M.;
RT "Differential expression of four members of the POU family of
RT proteins in activated and phorbol 12-myristate 13-acetate-treated
RT Jurkat T cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10260-10264(1993).
RN [3]
RP SEQUENCE OF 85-423 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93027214; PubMed=1357630;
RA Collum R.G., Fisher P.E., Datta M., Mellis S., Thiele C., Huebner K.,
RA Croce C.M., Israel M.A., Theil T., Moroy T., Depinho R.A., Alt F.W.;
RT "A novel POU homeodomain gene specifically expressed in cells of the
RT developing mammalian nervous system";
RL Nucleic Acids Res. 20:4919-4925(1992).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN
CC THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF
CC NEURONAL LINEAGES. MAY PLAY A ROLE IN DETERMINING OR MAINTAINING
CC THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: BRAIN. SEEMS TO BE SPECIFIC TO THE RETINA.
CC PRESENT IN THE DEVELOPING BRAIN, SPINAL CORD AND EYE.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS EARLY IN EMBRYOGENESIS
CC (DAY 13.5) AND IS UNDETECTABLE 14 DAYS AFTER BIRTH.
CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.
CC CLASS-4 SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U10063; AAA57161.1;
CC EMBL; U10062; AAA57161.1; JOINED.
DB DR
```

```
EMBL; L20433; AAA65605.1;
EMBL; X64624; CAA45907.1;
HSP: P10037; LAU7
TRANSFAC; T01876;
TRANSFAC; T04463;
Genew; HGNC:9218; POU4F1.
MIM; 601632;
InterPro; IPR001356; Homeobox.
InterPro; IPR000327; POU_domain.
Pfam; PF00046; homeobox; 1.
PRINTS; PR00028; POU_DOMAIN.
ProDom; PD000010; Homeobox; 1.
ProDom; PD000583; POU_domain; 1.
SMART; SM00389; HOX; 1.
SMART; SM00352; POU; 1.
PROSITE; PS00027; HOMEBOX_1; 1.
PROSITE; PS00035; POU_1; 1.
PROSITE; PS00465; POU_2; 1.
PROSITE; PS0071; HOMEBOX_2; 1.
Transcription regulation; Nuclear protein; DNA-binding; Homeobox;
KW Developmental protein.
FT DOMAIN 57 66 POU-IV BOX.
FT DOMAIN 100 108 POLY-HIS.
FT DOMAIN 130 189 GLY-RICH.
FT DOMAIN 207 260 ALA-RICH.
FT DOMAIN 267 341 POU.
FT DNA_BIND 359 418 HOMEBOX.
FT CONFLICT 91 92 TS -> H (IN REF. 3).
FT CONFLICT 99 99 A -> R (IN REF. 1).
FT CONFLICT 133 135 GAG -> ARR (IN REF. 3).
FT CONFLICT 135 137 MISSING (IN REF. 2).
FT CONFLICT 139 142 MISSING (IN REF. 3).
FT CONFLICT 149 150 GP -> AA (IN REF. 3).
FT CONFLICT 155 156 GP -> PR (IN REF. 3).
FT CONFLICT 174 175 GP -> AA (IN REF. 3).
FT CONFLICT 189 189 G -> A (IN REF. 3).
FT CONFLICT 240 240 A -> S (IN REF. 2).
FT CONFLICT 252 252 MISSING (IN REF. 3).
FT CONFLICT 256 256 A -> R (IN REF. 3).
FT CONFLICT 276 278 AER -> GS (IN REF. 3).
FT CONFLICT 343 345 AQR -> PS (IN REF. 3).
SQ SEQUENCE 423 AA; 42939 MW; B3AE4732E1309F34 CRC64;

Query Match      29.2%; Score 90; DB 1; Length 423;
Best Local Similarity 60.0%; Pred. No. 0.15;
Matches 18; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GDAPAGGVPGGAVPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 37
DB 153 GGGPGGGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 182

RESULT 7
BR3A_MOUSE
ID AC P17208; STANDARD; PRT; 421 AA.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Brain-specific homeobox/POU domain protein 3A (BRN-3A) (BRN-3.0).
GN POU4F1 OR BRN3A OR BRN3 OR BRN-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9421519; PubMed=8162704;
RA Theil T., Zechner U., Klett C., Adolph S., Moercoy T.;
RT "Chromosomal localization and sequences of the murine Brn-3 family of
RT developmental control genes.";
RL Cytogenet. Cell Genet. 66:267-271(1994).
```

```

RN  [2]
RP  SEQUENCE OF 286-401 FROM N.A.
RC  STRAIN=T6 / TWI; TISSUE=Testis;
RA  MEDLINE=90221898; PubMed=1970171;
RX  Goldsborough A., Ashworth A., Willison K.;
RT  "Cloning and sequencing of POU-boxes expressed in mouse testis.";
RL  Nucleic Acids Res. 18:1634-1634(1990).
CC  -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN
CC  THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF
CC  NEURONAL LINEAGES. MAY PLAY A ROLE IN DETERMINING OR MAINTAINING
CC  THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
CC  -1- SUBCELLULAR LOCATION: Nuclear.
CC  -1- TISSUE SPECIFICITY: BRAIN, PERIPHERAL SENSORY NERVOUS SYSTEM AND
CC  RETINA. IN THE ADULT NERVOUS SYSTEM BRN-3.0 PREDOMINATES IN THE
CC  MEDIAL HABENULA, SUPERFICIAL GRAY OF THE SUPERIOR COLLICULUS, RED
CC  NUCLEUS, MESENCEPHALIC NUCLEUS OF THE TRIGEMINAL GANGLION, NUCLEUS
CC  AMBIGUUS, INFERIOR OLIVARY NUCLEUS, AND PERIPHERAL SENSORY
CC  GANGLIA.
CC  -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.
CC  CLASS-4 SUBFAMILY.
CC  -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; S69350; AAB30577.2;
CC  EMBL; X51959; CAA36218.1;
CC  PIR; S09237; S09237.
CC  HSP; P10037; IAU7.
CC  MGD; MGI:102525; Pou4f1.
CC  InterPro: IPR001356; Homeobox.
CC  InterPro: IPR000327; POU_domain.
CC  Pfam: PF00046; homeobox; 1.
CC  Pfam: PF00157; POU; 1.
CC  PRINTS; PR00028; POU_DOMAIN.
CC  ProDom; PD000010; Homeobox; 1.
CC  ProDom; PD000583; POU_domain; 1.
CC  SMART; SM00389; HOX; 1.
CC  SMART; SM00352; POU; 1.
CC  PROSITE; PS00027; HOMEBOX_1; 1.
CC  PROSITE; PS00035; POU_1; 1.
CC  PROSITE; PS00465; POU_2; 1.
CC  PROSITE; PS50071; HOMEBOX_2; 1.
CC  Transcription regulation; Nuclear protein; DNA-binding; Homeobox;
CC  Developmental protein.
CC  DOMAIN 57 66 POU-IV BOX.
CC  DOMAIN 100 108 POLY-HIS.
CC  DOMAIN 130 187 GLY-RICH.
CC  DOMAIN 205 258 ALA-RICH.
CC  DOMAIN 265 339 POU.
CC  DNA_BIND 357 416 HOMEBOX.
CC  SEQUENCE 421 AA; 42781 MW; 34EC99D789EBE939 CRC64;
Query Match 27.8%; Score 85.5; DB 1; Length 421;
Best Local Similarity 62.1%; Pred. No. 0.38;
Matches 18; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 8 GDAPAGGVPGGVPGGFDGPGGPGG 36
DB 150 GGGPGGGGPGGGPGG-GGGGGGPGGG 177
RESULT 8
HS7L_PICAN
ID HS7L_PICAN STANDARD; PRT; 644 AA.
AC P53421;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Heat-shock protein 70 1 (HSP72).
GN HSA1
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 34438;
RX MEDLINE=96437974; PubMed=8840502;
RA Titorenko V.I., Evers M.E., Diesel A., Samyn B., van Beeumen J.,
RA Roggenkamp R.O., Kiel J.A.K.W., van der Klei I.J., Veenhuis M.;
RT "Identification and characterization of cytosolic Hansenula
RT polymorpha proteins belonging to the Hsp70 protein family.";
RL Yeast 12:849-857(1996).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC  EMBL; 229379; CAA82570.1;
CC  HSP; P19120; 3HSC.
CC  InterPro: IPR001023; Hsp70.
CC  Pfam; PF00012; HSP70; 1.
CC  PRINTS; PR00301; HEATSHOCK70.
CC  ProDom; PD000089; Hsp70; 1.
CC  PROSITE; PS00297; HSP70_1; 1.
CC  PROSITE; PS00329; HSP70_2; 1.
CC  PROSITE; PS01036; HSP70_3; 1.
CC  Chaperone; Heat shock; ATP-binding; Multigene family; Acetylation.
CC  INIT_MET 0 0
CC  MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC  SEQUENCE 644 AA; 70006 MW; C5C908D9066501C4 CRC64;
Query Match 26.8%; Score 82.5; DB 1; Length 644;
Best Local Similarity 48.9%; Pred. No. 1;
Matches 22; Conservative 1; Mismatches 11; Indels 11; Gaps 3;
QY 9 DAPAGGVPGGVPGGFDGPGGPGG 53
DB 608 DLYAGGVPGGAAPGGFP-----GAGGNAP-----GAD-QGPSVE 641
RESULT 9
K1CJ_HUMAN
ID K1CJ_HUMAN STANDARD; PRT; 593 AA.
AC P13645;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (K10) (CK 10).
GN KRT10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89125611; PubMed=2464696;
RA Rieger M., Franke W.W.;
RT "Identification of an orthologous mammalian cytokeatin gene. High
RT degree of intron sequence conservation during evolution of human
RT cytokeatin 10.";
RL J. Mol. Biol. 204:841-856(1988).

```


FT DOMAIN 493 533 EGF-LIKE 6.
FT DOMAIN 534 573 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 574 615 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 616 656 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 657 697 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 698 766 TGFBP 2.
FT DOMAIN 767 808 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 809 850 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 851 890 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 891 935 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 936 1071 TGFBP 3.
FT DOMAIN 1072 1113 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1114 1156 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1157 1198 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1199 1240 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1241 1281 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1282 1323 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1324 1365 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1366 1406 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1407 1447 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1448 1489 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1490 1530 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1531 1571 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1572 1648 TGFBP 4.
FT DOMAIN 1649 1690 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1691 1732 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1733 1806 TGFBP 5.
FT DOMAIN 1807 1848 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1891 1932 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1933 1971 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1972 2014 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2015 2054 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2055 2096 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2097 2169 TGFBP 6.
FT DOMAIN 2170 2211 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2212 2251 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2252 2292 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2293 2336 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2337 2378 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2379 2447 TGFBP 7.
FT DOMAIN 2448 2489 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2490 2530 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2531 2569 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2570 2612 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2613 2652 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2653 2693 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2694 2733 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 119 130 BY SIMILARITY.
FT DISULFID 132 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 195 BY SIMILARITY.
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 209 221 BY SIMILARITY.
FT DISULFID 226 300 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 321 333 BY SIMILARITY.
FT DISULFID 328 342 BY SIMILARITY.
FT DISULFID 344 357 BY SIMILARITY.
FT DISULFID 497 509 BY SIMILARITY.
FT DISULFID 504 518 BY SIMILARITY.
FT DISULFID 520 532 BY SIMILARITY.
FT DISULFID 538 548 BY SIMILARITY.
FT DISULFID 543 557 BY SIMILARITY.
FT DISULFID 559 572 BY SIMILARITY.
FT DISULFID 578 590 BY SIMILARITY.
FT DISULFID 585 599 BY SIMILARITY.
FT DISULFID 601 614 BY SIMILARITY.
FT DISULFID 620 631 BY SIMILARITY.

FT DISULFID 626 640 BY SIMILARITY.
FT DISULFID 642 655 BY SIMILARITY.
FT DISULFID 661 672 BY SIMILARITY.
FT DISULFID 667 681 BY SIMILARITY.
FT DISULFID 683 696 BY SIMILARITY.
FT DISULFID 771 783 BY SIMILARITY.
FT DISULFID 778 792 BY SIMILARITY.
FT DISULFID 794 807 BY SIMILARITY.
FT DISULFID 813 825 BY SIMILARITY.
FT DISULFID 820 834 BY SIMILARITY.
FT DISULFID 836 849 BY SIMILARITY.
FT DISULFID 855 865 BY SIMILARITY.
FT DISULFID 860 874 BY SIMILARITY.
FT DISULFID 876 889 BY SIMILARITY.
FT DISULFID 958 970 BY SIMILARITY.
FT DISULFID 965 979 BY SIMILARITY.
FT DISULFID 981 994 BY SIMILARITY.
FT DISULFID 1076 1088 BY SIMILARITY.
FT DISULFID 1083 1097 BY SIMILARITY.
FT DISULFID 1099 1112 BY SIMILARITY.
FT DISULFID 1118 1130 BY SIMILARITY.
FT DISULFID 1125 1139 BY SIMILARITY.
FT DISULFID 1141 1155 BY SIMILARITY.
FT DISULFID 1161 1173 BY SIMILARITY.
FT DISULFID 1168 1182 BY SIMILARITY.
FT DISULFID 1184 1197 BY SIMILARITY.
FT DISULFID 1203 1215 BY SIMILARITY.
FT DISULFID 1210 1224 BY SIMILARITY.

Query Match 26.68; Score 82; DB 1; Length 2911;
Best Local Similarity 35.58; Pred. No. 4.3;
Matches 22; Conservative 10; Mismatches 14; Indels 16; Gaps 5;

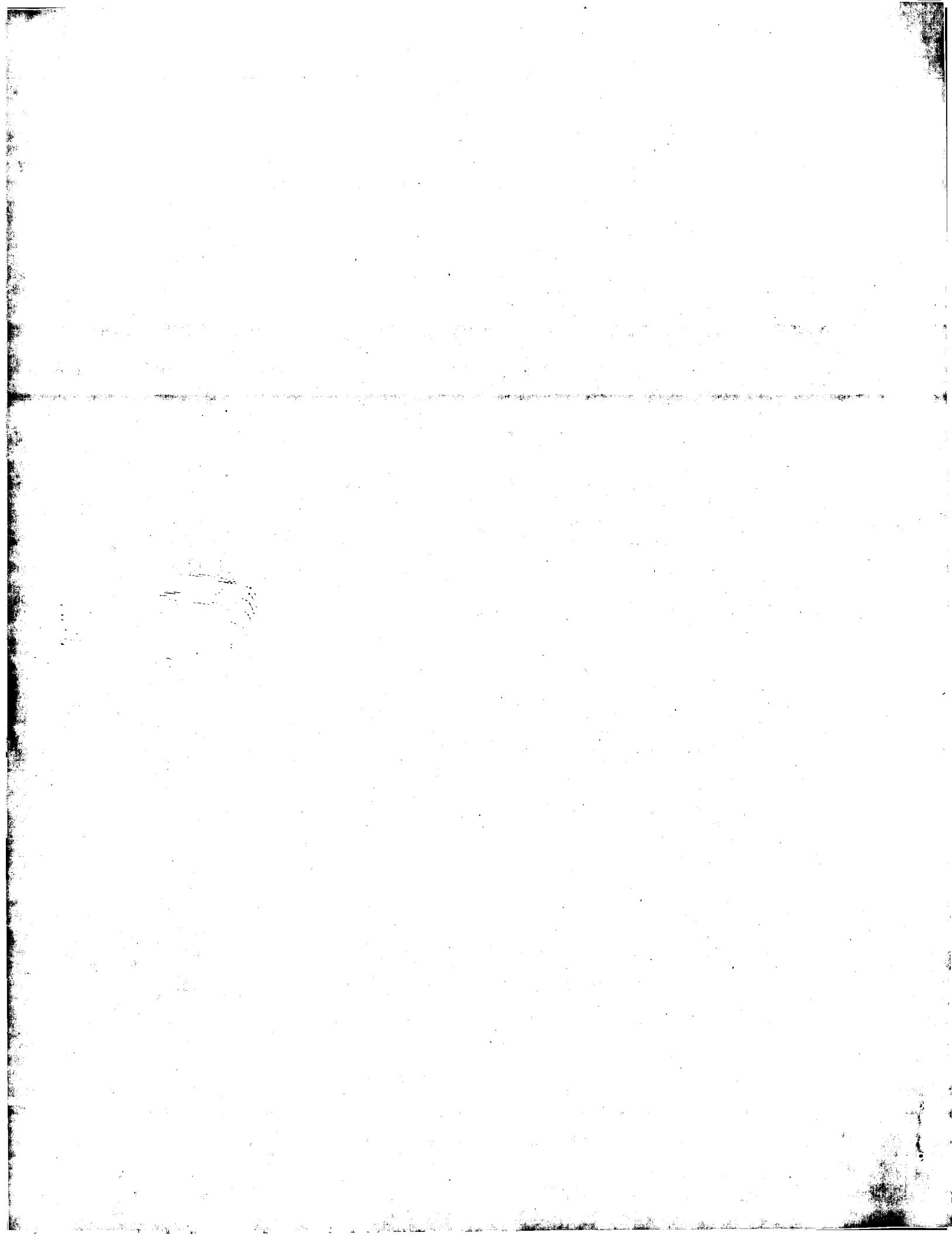
Oy 8 GDAPAGG---VPGA---VPG---GFGPGFVLDGMYGVDSSTVEL 54
Db 440 GFASGNGYGPCTGPIPIGGNGFSPGVGAGVGAGGOGPIIT---GLTILNQITDI 496
Oy 55 AQ 56
Db 497 CK 498

RESULT 11
DDX9_MOUSE STANDARD; PRT; 1380 AA.
ID DDX9_MOUSE
AC 070133; O35931; O54703;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent RNA helicase A (Nuclear DNA helicase II) (NDH II)
DE (DEAD-box protein 9) (mHEL-5).
GN DDX9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98149984; PubMed=9480750;
RA Lee C.-G., Eki T., Okumura K., da Costa Soares V., Hurwitz J.;
RT "Molecular analysis of the cDNA and genomic DNA encoding mouse RNA
helicase A." Genomics 47:365-371(1998).
RN [2]
RP SEQUENCE OF 386-919 FROM N.A.
RC STRAIN=C57BL/6;
RA Kiselow P., Miasek A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UNWINDS DOUBLE-STRANDED DNA AND RNA IN A 3' TO 5'
CC DIRECTION. GENERATES MULTIPLE SECONDARY STRUCTURES THAT INFLUENCE
CC RNA-BINDING PROTEINS. MAY PLAY A ROLE IN X-LINKED GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.


```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M00978; AAA29625.1; -.
CC PIR; A49242; A49242.
CC HSP; P08109; ICKR.
CC InterPro; IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC PRODOM; PD000089; Hsp70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC KW ATP-binding; Heat shock; Repeat.
CC DOMAIN 632 675 10 X 4 AA TANDEM REPEATS OF G-G-N-P.
CC FT REPEAT 632 635 1.
CC FT REPEAT 636 639 2.
CC FT REPEAT 640 643 3.
CC FT REPEAT 644 647 4.
CC FT REPEAT 648 651 5.
CC FT REPEAT 653 656 6.
CC FT REPEAT 657 660 7.
CC FT REPEAT 663 666 8.
CC FT REPEAT 668 671 9.
CC FT REPEAT 672 675 10.
CC SQ SEQUENCE 686 AA; 74602 MW; FF31F448FFCBF286 CRC64;
Query Match 46.3%; Score 81; DB 1; Length 686;
Best Local Similarity 43.4%; Pred No. 1.5;
Matches 23; Conservative 3; Mismatches 11; Indels 16; Gaps 4;
QY 8 GDAPAG--GGVPGGAVPGFD-----PGFGPGGFGVLGDGYGVDSGSTVE 53
Db 1 ||| ||| :||| : ||| ||| ||| ||| ||| |||
640 GGMPGGMPGGMPGGMPGMNFGMGPMGGMPGG-APA-----GSGPTVE 683
RESULT 13
HS70_BLAEM STANDARD; PRT; 649 AA.
ID HS70_BLAEM AC P48720;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Heat shock 70 kDa protein.
GN HSP70.
OS Blastocladiaella emersonii.
OC Eukaryota; Fungi; Chytridiomycota; Blastocladales; Blastocladiaceae;
OC Blastocladiaella.
ON NCBI_TaxID=4308;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95123910; PubMed=7828923;
RX Stefani R.M.; Gomes S.L.;
RT "A unique intron-containing hsp70 gene induced by heat shock and
RT during sporulation in the aquatic fungus Blastocladiaella emersonii.";
RL Gene 152:19-26(1995).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L22497; AAA65099.1; -.
CC HSSP; P19120; 3HSC.
CC InterPro; IPR001023; Hsp70.
DR
```


FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 FT CONFLICT 175 175 A -> G (IN REF. 2).
 SQ SEQUENCE 655 AA; 70192 MW; 634743E4D6DAD9D5 CRC64;
 Query Match 26.0%; Score 80; DB 1; Length 655;
 Best Local Similarity 50.0%; Pred. No. 1.8;
 Matches 24; Conservative 1; Mismatches 13; Indels 10; Gaps 4;
 Qy 8 GDAPAG-GGVPG-GAVPGGFDGPGGPGGPGVLDGWYGVDSGSTVE 53
 Db 613 GGAPGGAGGPGGPGG-APGAGGPGGA-----TGGESSGPTVE 652

Search completed: May 7, 2003, 16:48:04
 Job time : 7.43782 secs



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	308	100.0	160	2	Q9K5G3	Q9K5G3 bordetella
2	308	100.0	161	2	Q9K5G2	Q9K5G2 bordetella
3	301	97.7	159	2	Q9ALP4	Q9ALP4 bordetella
4	291	94.5	910	2	Q69259	Q69259 bordetella
5	290.5	94.3	184	2	Q9ALP6	Q9ALP6 bordetella
6	290.5	94.3	195	2	Q9ALP8	Q9ALP8 bordetella
7	287	93.2	907	2	Q69257	Q69257 bordetella
8	281	91.2	387	2	Q9S3M9	Q9S3M9 bordetella
9	281	91.2	910	2	Q9S6N1	Q9S6N1 bordetella
10	278.5	90.4	915	2	Q88143	Q88143 bordetella
11	276	89.6	851	2	Q9ALX8	Q9ALX8 bordetella
12	276	89.6	920	2	Q93L98	Q93L98 bordetella
13	274.5	89.1	912	2	Q8RSU0	Q8RSU0 bordetella
14	268.5	87.2	167	2	Q937U5	Q937u5 bordetella
15	260.5	84.6	181	2	Q9ALQ1	Q9ALq1 bordetella
16	255.5	83.0	905	2	Q9S3M8	Q9S3m8 bordetella

```
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Pertactin (P.68) (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEI;
RX MEDLINE=20359389; PubMed=1089896;
RA Boursaux-Eude C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250094; CAB76448.1; -
FT NON_TER 1
FT SEQUENCE 161 AA; 15435 MW; 7FACB18F1FD0507E CRC64;
SQ
Query Match 100.0%; Score 308; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 7.5e-24;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 56
|||||
DB 54 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 109
|||||

RESULT 3
Q9ALP4 PRELIMINARY; PRT; 159 AA.
AC Q9ALP4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV6;
RX MEDLINE=21117018; PubMed=11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella bronchiseptica Pertactin.";
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL; AY007276; AAG38452.1; -
DR InterPro; IPR003992; pertactin.
DR PRINTS; PR01482; PERTACTIN.
FT NON_TER 1
FT NON_TER 159
FT SEQUENCE 159 AA; 15230 MW; 647F0CA00634915F CRC64;
SQ
Query Match 97.7%; Score 301; DB 2; Length 159;
Best Local Similarity 98.2%; Pred. No. 3.8e-23;
Matches 55; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 56
|||||
DB 79 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 134
|||||

RESULT 4
O69259 PRELIMINARY; PRT; 910 AA.
ID O69259
AC O69259;
DT 01-AUG-1998 (Tremblrel. 07, Created)
```

```
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Pertactin precursor (Pertactin outer membrane protein).
GN PRN OR PRN3 OR PRNA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=287, AND AL1561;
RX MEDLINE=99345256; PubMed=10418915;
RA Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;
RT "Intranasal murine model of Bordetella pertussis infection. II. Sequence variation and protection induced by a tricomponent acellular vaccine.";
RL Vaccine 17:2651-2660(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B343;
RX MEDLINE=98114370; PubMed=9453625;
RA Moel F.R., VanOirschot H., Heuvelman K., vanderHeide H., Gastra W., Willems R.R.J.;
RT "Polymorphism in the Bordetella pertussis virulence factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution.";
RL Infect. Immun. 66:670-675(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=99R45;
RX MEDLINE=21391812; PubMed=11500425;
RA Fry N.K., Neal S., Harrison T.G., Miller E., Matthews R., George R.C.;
RT "Genotypic Variation in the Bordetella pertussis Virulence Factors Pertactin and Pertussis Toxin in Historical and Recent Clinical Isolates in the United Kingdom.";
RL Infect. Immun. 69:5520-5528(2001).
DR EMBL; AJ006156; CAA06898.2; -
DR EMBL; AJ011093; CAA09475.1; -
DR EMBL; AJ007362; CAA07477.1; -
DR EMBL; AF348485; AAK92093.1; -
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMLY.
KW Signal.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 910 PERTACTIN.
FT SEQUENCE 910 AA; 93486 MW; 9035EE1F07A52BC8 CRC64;
SQ
Query Match 94.5%; Score 291; DB 2; Length 910;
Best Local Similarity 94.6%; Pred. No. 2.2e-21;
Matches 53; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 56
|||||
DB 254 QRATIRGDAPAGGVPGGVPGGFGPGGFGPVLGDWGVGVDSGSTVELAQ 309
|||||

RESULT 5
Q9ALP6 PRELIMINARY; PRT; 184 AA.
ID Q9ALP6;
AC Q9ALP6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
```

DT	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE	Pertactin precursor.	
GN	PRN OR PRN6.	
OS	Bordetella pertussis.	
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;	
OC	Bordetella.	
OX	NCBI_TaxID=520;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=18323;	
RA	Boursaux-Eude C., Guiso N.;	
RL	Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=18323;	
RA	van Oirschot H.F.L.M., Mooi F.R.;	
RL	Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CZ;	
RX	MEDLINE=99345256; PubMed=10418915;	
RA	Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;	
RT	"Intranasal murine model of Bordetella pertussis infection: II.	
RT	Sequence variation and protection induced by a tricomponent acellular	
RT	vaccine.";	
RL	Vaccine 17:2651-2660(1999).	
DR	EMBL; AJ006152; CAU06894.1; -.	
DR	EMBL; AJ132095; CAA10566.1; -.	
DR	EMBL; AJ006160; CAA06902.2; -.	
DR	InterPro; IPR003992; Pertactin.	
DR	InterPro; IPR003991; Pertactin_vir.	
DR	InterPro; IPR004899; Pertact_sup.	
DR	Pfam; PF03212; Pertactin; 1.	
DR	PRINTS; PRO1482; PERTACTIN.	
DR	PRINTS; PRO1484; PERTACTNFAMLY.	
KW	Signal.	
FT	SIGNAL	1 34 POTENTIAL.
FT	CHAIN	35 907 PERTACTIN.
SQ	SEQUENCE	907 AA; 93298 MW; 3A7D05F4094420EA CRC64;
Query Match 93.2%; Score 287; DB 2; Length 907;		
Best Local Similarity 94.6%; Pred. No. 5.6e-21;		
Matches 53; Conservative 1; Mismatches 2; Indels 0; Gaps		
QY	1	QRATIRRGDAPAGGVPGGAVPGDFPGCGFGPVLGDGWYG 56
Dd	254	QRATIRRGDAPAGGVPGGAVPGDFPGCGFGPVLGDGWYG 309
RESULT 8		
ID	Q9S3M9	PRELIMINARY; PRT: 387 AA.
AC	Q9S3M9;	
DT	01-MAY-2000	(TREMBLrel. 13, Created)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)
DE	Pertactin precursor (Fragment).	
GN	PRN.	
OS	Bordetella pertussis.	
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;	
OC	Bordetella.	
OX	NCBI_Taxid=520;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=W28;	
RX	MEDLINE=99345256; PubMed=10418915;	
RA	Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;	
RT	"Intranasal murine model of Bordetella pertussis infection: II.	
RT	Sequence variation and protection induced by a tricomponent acellular	
RT	vaccine.";	
RL	Vaccine 17:2651-2660(1999).	

DT	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE	Pertactin precursor.	
GN	PRN OR PRN6.	
OS	Bordetella pertussis.	
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;	
OC	Bordetella.	
OX	NCBI_TaxID=520;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=18323;	
RA	Boursaux-Eude C., Guiso N.;	
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=18323;	
RA	van Oirschot H.F.L.M., Mooi F.R.;	
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CZ;	
RX	MEDLINE=99345256; PubMed=10418915;	
RA	Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;	
RT	"Intranasal murine model of Bordetella pertussis infection: II.	
RT	Sequence variation and protection induced by a tricomponent acellular	
RT	vaccine.";	
RL	Vaccine 17:2651-2660(1999).	
DR	EMBL; AJ006152; CAA06894.1; -.	
DR	EMBL; AJ132095; CAA10566.1; -.	
DR	EMBL; AJ006160; CAA06902.2; -.	
DR	InterPro; IPR003992; pertactin.	
DR	InterPro; IPR003991; pertactin_vir.	
DR	InterPro; IPR004899; Pertactin_sup.	
DR	Pfam; PF03212; Pertactin; 1.	
DR	PRINTS; PR01482; PERTACTIN.	
DR	PRINTS; PR01484; PERTACTNFAMLY.	
KW	Signal.	
FT	SIGNAL	1 34 POTENTIAL.
FT	CHAIN	35 907 PERTACTIN.
SQ	SEQUENCE	907 AA; 93298 MW; 3A7D05F4094420EA CRC64;
Query Match 93.2%; Score 287; DB 2; Length 907;		
Best Local Similarity 94.6%; Pred. No. 5.6e-21;		
Matches 53; Conservative 1; Mismatches 2; Indels 0; Gaps		
QY	1	QRATIRRGDAPAGGVPGGAVPGGFGPVGCGFGPVLGDWGYGV 56
DB	254	QRATIRRGDAPAGGVPGGAVPGGFGPVGCGFGPVLGDWGYGV 309
RESULT 8		
ID	Q9S3M9	PRELIMINARY; PRT: 387 AA.
AC	Q9S3M9;	
DT	01-MAY-2000	(TREMBLrel. 13, Created)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)
DE	Pertactin precursor (Fragment).	
GN	PRN.	
OS	Bordetella pertussis.	
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;	
OC	Bordetella.	
OX	NCBI_TaxID=520;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=W28;	
RX	MEDLINE=99345256; PubMed=10418915;	
RA	Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;	
RT	"Intranasal murine model of Bordetella pertussis infection: II.	
RT	Sequence variation and protection induced by a tricomponent acellular	
RT	vaccine.";	
RL	Vaccine 17:2651-2660(1999).	

```
DR EMBL; AJ006154; CAA06896.1; -.
KW Signal. 1 34 POTENTIAL.
FT SIGNAL. 35 >387 PERTACTIN.
FT CHAIN 387
FT NON_TER 387
SQ SEQUENCE 387 AA; 38714 MW; 61195D45D347ALEB CRC64;

Query Match 91.2%; Score 281; DB 2; Length 387;
Best Local Similarity 92.9%; Pred. No. 9.6e-21;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGVPGGFPGGFGPVLGDWGYGVDSGSTVELAQ 56
Db 254 QRATIRGDAPAGGAVPGGVPGGFPGGFGPVLGDWGYGVDSGSTVELAQ 309

RESULT 9
Q9S6N1 PRELIMINARY; PRT; 910 AA.
AC Q9S6N1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE Pertactin.
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN PRN7.
OC Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B567;
RA van Oirschot H.F.I.M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-B567;
RX MEDLINE-98114370; PubMed-9453625;
RA Mool F.R., VanOirschot H., Heuvelman K., vanderHeide H., Gaastra W.,
  Willem's R.J.J.;
RT "Polymorphism in the Bordetella pertussis virulence factors
  P.69/pertactin and pertussis toxin in The Netherlands: temporal trends
  and evidence for vaccine-driven evolution.";
RL Infect. Immun. 66:670-675(1998).
DR EMBL; AJ133784; CAB40080.1; -.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR Pfam; PF03212; Pertactin.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PRTECTNFAMLY.
SQ SEQUENCE 910 AA; 93495 MW; AF38246F8D82E03D CRC64;

Query Match 91.2%; Score 281; DB 2; Length 910;
Best Local Similarity 92.9%; Pred. No. 2.3e-20;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGVPGGFPGGFGPVLGDWGYGVDSGSTVELAQ 56
Db 254 QRATIRGDAPAGGAVPGGVPGGFPGGFGPVLGDWGYGVDSGSTVELAQ 309

RESULT 10
O88143 PRELIMINARY; PRT; 915 AA.
AC O88143;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE P.69B protein precursor (Pertactin outer membrane protein).
GN PRN2 OR PRNA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
```

```
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B345;
RX MEDLINE-98114370; PubMed-9453625;
RA Mool F.R., VanOirschot H., Heuvelman K., Vanderheide H.G., Gaastra W.,
  Willem's R.J.J.;
RT "Polymorphism in the Bordetella pertussis virulence factors
  P.69/pertactin and pertussis toxin in The Netherlands: temporal trends
  and evidence for vaccine-driven evolution.";
RL Infect. Immun. 66:670-675(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HAV;
RA Boursaux-Eude C., Thiberge S., Carletti G., Guiso N.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-98K320, AND DCH132;
RX MEDLINE-21391812; PubMed-11500425;
RA Fry N.K., Neal S., Harrison T.G., Miller E., Matthews R., George R.C.;
RT "Genotypic Variation in the Bordetella pertussis Virulence Factors
  pertactin and Pertussis Toxin in Historical and Recent Clinical
  Isolates in the United Kingdom.";
RL Infect. Immun. 69:5520-5528(2001).
DR EMBL; AJ011092; CAA09474.1; -.
DR EMBL; AJ007361; CAA07476.1; -.
DR EMBL; AF348484; AAK92092.1; -.
DR EMBL; AF348482; AAK92090.1; -.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PRTECTNFAMLY.
KW Signal. 1 34 POTENTIAL.
FT SIGNAL. 35 915 P.69B PROTEIN.
FT CHAIN 35
SQ SEQUENCE 915 AA; 93902 MW; 4AF56A29BAAFFED0 CRC64;

Query Match 90.4%; Score 278.5; DB 2; Length 915;
Best Local Similarity 86.9%; Pred. No. 4.1e-20;
Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGVPGGAV-----PGGFDPGGFGPGGFGPVLGDWGYGVDSGSTVELA 55
Db 254 QRATIRGDAPAGGVPGGAVPGGFGPGGFGPVLGDWGYGVDSGSTVELA 313

QY 56 Q 56
Db 314 Q 314

RESULT 11
Q9AIX8 PRELIMINARY; PRT; 851 AA.
AC Q9AIX8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pertactin outer membrane protein (Fragment).
GN PRNA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MO-908;
RA Fry N.K., Neal S., Li Y.T., Hanauer S., Blank E., Harrison T.G.;
RT "Sequence variation in the pertactin and pertussis toxin genes and
  amplified fragment length polymorphism analysis of Bordetella
```


RT pertussis isolates from Missouri, USA.*;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218785; AAK31810.1; -
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vlr.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMLY.
FT NON_TER 851 851
FT NON_TER 851 851
SQ SEQUENCE 851 AA; 86636 MW; B2D08B3519BDBDC CRC64;

Query Match 89.68; Score 276; DB 2; Length 851;
Best Local Similarity 80.3%; Pred. No. 6.8e-20;
Matches 53; Conservative 1; Mismatches 2; Indels 10; Gaps 1;
QY 1 QRATIRRGDAPAGGVPGGAV-----PGGFDGPGGFGPGVLDGWYGVDSGS 50
|||||
Db 245 QRATIRRGDAPAGGVPGGAVPGGFGPGGFGPGGFGPGGFGPGVLDGWYGVDSGS 304
|||||
QY 51 TVELAQ 56
:|||||
Db 305 SVELAQ 310

RESULT 12

Q93L98 ID Q93L98 PRELIMINARY; PRT; 920 AA.
AC Q93L98;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Pertactin.
GN PRN9.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B1679;
RA Heuvelman K., Peppier M.S., Lewandowski A., Mooi F.R.;
RT "Bordetella pertussis prn gene for pertactin.*";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ315611; CAC42396.1; -
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR Pfam; PF03212; Pertactin; 1.
SQ SEQUENCE 920 AA; 94317 MW; 62AA9461DCF53E54 CRC64;

Query Match 89.68; Score 276; DB 2; Length 920;
Best Local Similarity 80.3%; Pred. No. 7.4e-20;
Matches 53; Conservative 1; Mismatches 2; Indels 10; Gaps 1;
QY 1 QRATIRRGDAPAGGVPGGAV-----PGGFDGPGGFGPGVLDGWYGVDSGS 50
|||||
Db 254 QRATIRRGDAPAGGVPGGAVPGGFGPGGFGPGGFGPGGFGPGVLDGWYGVDSGS 313
|||||
QY 51 TVELAQ 56
:|||||
Db 314 SVELAQ 319

RESULT 13

Q8RSU0 ID Q8RSU0 PRELIMINARY; PRT; 912 AA.
AC Q8RSU0;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Pertactin.
GN PRN.
OS Bordetella pertussis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B442;
RA van Loo I.H., Mooi F.R.;
RT "Changes in the Bordetella pertussis population in the first 20 years
after the introduction of vaccination.*";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430832; CAD23609.1; -
SQ SEQUENCE 912 AA; 93680 MW; 7F417BA66B732EF7 CRC64;

Query Match 89.18; Score 274.5; DB 2; Length 912;
Best Local Similarity 86.9%; Pred. No. 1e-19;
Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;
QY 1 QRATIRRGDAPAGG-----YPGGAVPGGFGPGGFGPGVLDGWYGVDSGSSTVELA 55
|||||
Db 254 QRATIRRGDAPAGGVPGGAVPGGAVPGGFGPGGFGPGVLDGWYGVDSGSSTVELA 313
|||||
QY 56 Q 56
:|||||
Db 314 Q 314

RESULT 14

Q937U5 ID Q937U5 PRELIMINARY; PRT; 167 AA.
AC Q937U5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A67;
RA Weber C., Boursaux-Eude C., Nicole G.;
RT "Polymorphism of Bordetella pertussis isolates circulating the last
ten years in France, a country using the same effective whole-cell
vaccine since more than thirty years.*";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ318056; CAC67458.1; -
FT NON_TER 167 167
FT NON_TER 167 167
SQ SEQUENCE 167 AA; 15826 MW; E136B4CF809565F0 CRC64;

Query Match 87.28; Score 268.5; DB 2; Length 167;
Best Local Similarity 85.28; Pred. No. 7.5e-20;
Matches 52; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
QY 1 QRATIRRGDAPAGGVPGGAVPGGFGPGGFGPGVLDGWYGVDSGSSTVELA 55
|||||
Db 54 QRATIRRGDAPAGGVPGGAVPGGAVPGGFGPGGFGPGVLDGWYGVDSGSSTVELA 113
|||||
QY 56 Q 56
:|||||
Db 114 Q 114

RESULT 15

Q9ALQ1 ID Q9ALQ1 PRELIMINARY; PRT; 181 AA.
AC Q9ALQ1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Pertactin (Fragment).

Search completed: May 7, 2003, 16:53:10
Job time : 29.6201 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	298	100.0	52	22	US-09-855-754-17
2	284.5	95.5	52	22	US-09-855-754-16
3	282	94.6	56	22	US-09-855-754-18
4	271	90.9	52	22	US-09-855-754-15
5	271	90.9	911	21	US-09-791-537-37413
6	271	90.9	911	21	US-09-791-537-85698
					Sequence 17, Appl
					Sequence 16, Appl
					Sequence 18, Appl
					Sequence 15, Appl
					Sequence 37413, A
					Sequence 85698, A

Qy

14, Application NO. 335337/34

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	298	100.0	52	5	US-09-855-754B-17	Sequence 17, Appl	
2	284.5	95.5	59	5	US-09-855-754B-16	Sequence 16, Appl	
3	282	94.6	56	5	US-09-855-754B-18	Sequence 18, Appl	
4	271	90.9	52	5	US-09-855-754B-15	Sequence 15, Appl	
5	271	90.9	911	5	US-09-855-754B-4	Sequence 4, Appl	
6	271	90.9	911	6	US-10-227-353-4	Sequence 4, Appl	
7	267	89.6	58	5	US-09-855-754B-19	Sequence 19, Appl	
8	267	89.6	922	5	US-09-855-754B-6	Sequence 6, Appl	
9	267	89.6	922	6	US-10-227-353-6	Sequence 6, Appl	
10	250.5	84.1	49	5	US-09-855-754B-14	Sequence 14, Appl	
11	244	81.9	48	5	US-09-855-754B-20	Sequence 20, Appl	
12	236	79.2	52	5	US-09-855-754B-21	Sequence 21, Appl	
13	235	78.9	54	5	US-09-855-754B-22	Sequence 22, Appl	
14	207	69.5	42	5	US-09-855-754B-23	Sequence 23, Appl	
15	207	69.5	768	6	US-10-282-122A-51012	Sequence 51012, A	
16	207	69.5	910	5	US-09-855-754B-5	Sequence 5, Appl	
17	207	69.5	910	6	US-10-227-353-2	Sequence 2, Appl	
18	186.5	62.6	39	5	US-09-855-754B-24	Sequence 24, Appl	
19	140	47.0	242	1	PCR-US02-36123-2812	Sequence 2812, Ap	
20	140	47.0	355	1	PCR-US02-36123-2814	Sequence 2814, Ap	
21	139	46.6	523	6	US-10-017-161-1982	Sequence 1982, Ap	
22	138	45.3	270	6	US-10-282-122A-66157	Sequence 66157, A	
23	136	45.6	188	6	US-10-366-683-28564	Sequence 28564, A	
24	136	45.6	188	6	US-10-419-128-28564	Sequence 28564, A	
25	135	45.3	487	6	US-10-224-999A-3465	Sequence 3465, Ap	
26	133.5	44.8	900	6	US-10-218-140-4170	Sequence 4170, Ap	

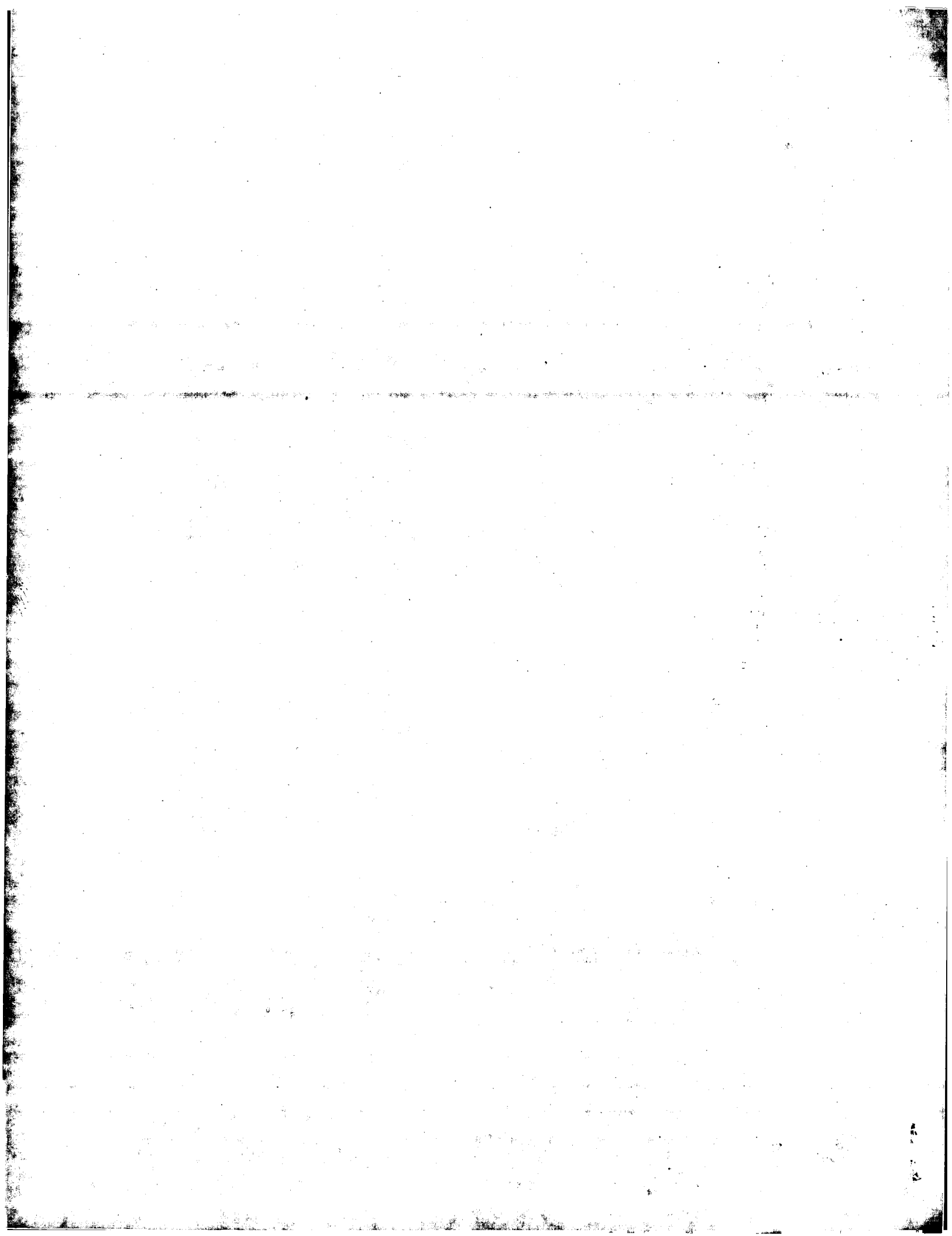

```

: FILE REFERENCE: 03495-0206-00000
: CURRENT APPLICATION NUMBER: US/09/855,754B
: CURRENT FILING DATE: 2001-05-16
: PRIOR APPLICATION NUMBER: 60/206,969
: PRIOR FILING DATE: 2000-03-25
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn Ver. 2.1

```

APPLICANT: GUIISO-MACLOUF, NIG

; APPLICANT: BOURSAUX-EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE



[illegible]

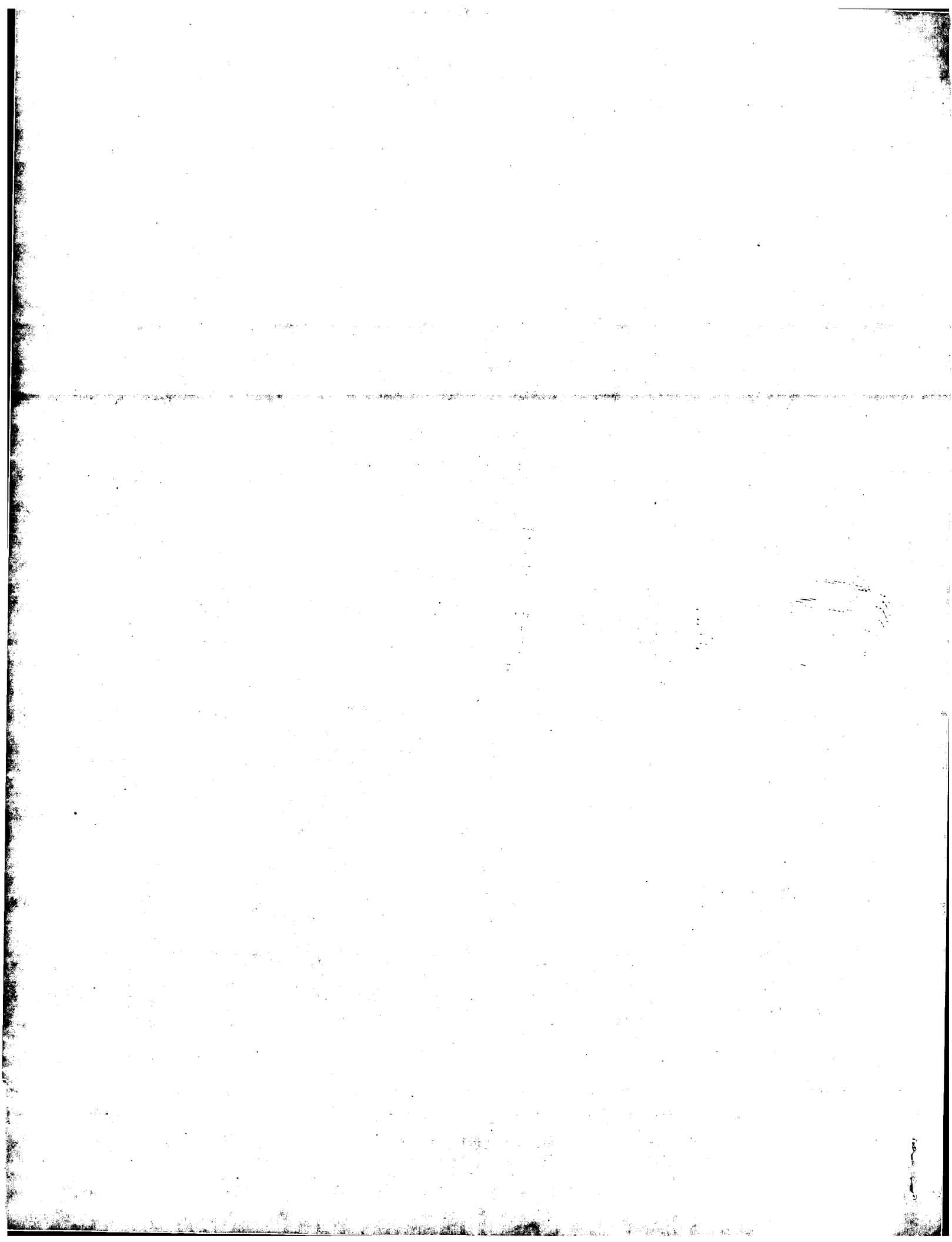
US-09-855-754-14
: Sequence 14. Application US/09855754

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	322	100.0	56	5	US-09-855-754B-18	Sequence 18, Appl	
2	303	94.1	58	5	US-09-855-754B-19	Sequence 19, Appl	
3	303	94.1	922	5	US-09-855-754B-6	Sequence 6, Appl	
4	303	94.1	922	6	US-10-227-353-6	Sequence 6, Appl	
5	294.5	91.5	59	5	US-09-855-754B-16	Sequence 16, Appl	
6	285	88.5	52	5	US-09-855-754B-15	Sequence 15, Appl	
7	285	88.5	911	5	US-09-855-754B-4	Sequence 4, Appl	
8	285	88.5	911	6	US-10-227-353-4	Sequence 4, Appl	
9	282	87.6	52	5	US-09-855-754B-17	Sequence 17, Appl	
10	264.5	82.1	49	5	US-09-855-754B-14	Sequence 14, Appl	
11	258	80.1	48	5	US-09-855-754B-21	Sequence 21, Appl	
12	257	79.8	52	5	US-09-855-754B-20	Sequence 20, Appl	
13	256	79.5	54	5	US-09-855-754B-22	Sequence 22, Appl	
14	210	65.2	42	5	US-09-855-754B-23	Sequence 23, Appl	
15	210	65.2	768	6	US-10-282-122A-51012	Sequence 51012, A	
16	210	65.2	910	5	US-09-855-754B-5	Sequence 5, Appl	
17	210	65.2	910	6	US-10-227-353-2	Sequence 2, Appl	
18	189.5	48.9	39	5	US-09-855-754B-24	Sequence 24, Appl	
19	155.5	48.3	900	6	US-10-218-140-4170	Sequence 4170, Ap	
20	155.5	48.3	1218	7	US-60-452-680-19541	Sequence 19541, A	
21	155.5	48.3	1218	7	US-60-453-135-11684	Sequence 11684, A	
22	155.5	48.3	1218	7	US-60-453-050-11684	Sequence 11684, A	
23	151	46.9	523	6	US-10-017-161-1982	Sequence 1982, Ap	
24	144	44.7	632	6	US-10-282-122A-51472	Sequence 51472, A	
25	143	44.4	242	1	PCF-US02-36123-2812	Sequence 2812, Ap	
26	143	44.4	355	1	PCF-US02-36123-2814	Sequence 2814, Ap	


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
;
; US-10-227-353-4
;
; Query Match 88.5%; Score 285; DB 6; Length
; Best Local Similarity 92.9%; Pred. No. 1.3e-10;
; Matches 52; Conservative 0; Mismatches 0; Indels
;
; QY 1 GAKAPPAPAPQPGQPGQPGQPGQPGQPGQPGQPGQPEAPAPQPPAGG
;
; Db 559 GAKAPPAPAPPA----PQGPQPGQPGQPGQPGQPGQPGQPEAPAPQPPAGG
;
;
; RESULT 9
; US-09-855-754B-17
;
; Sequence 17, Application US/09855754B
;
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
;
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA P
;
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDET
;
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOS
;
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
;
; FILE REFERENCE: 03495-0206-00000
;
; CURRENT APPLICATION NUMBER: US/09/855,754B
;
; CURRENT FILING DATE: 2001-05-16
;
; PRIOR APPLICATION NUMBER: 60/206, 969
;
; PRIOR FILING DATE: 2000-05-25
;
; NUMBER OF SEQ ID NOS: 25
;
; SOFTWARE: Patent In Ver. 2.1
;
; SEQ ID NO 17
;
; LENGTH: 52
;
; TYPE: PRT
;
; ORGANISM: Bordetella bronchiseptica
;
; US-09-855-754B-17
;
; Query Match 87.6%; Score 282; DB 5; Length
; Best Local Similarity 91.1%; Pred. No. 3.4e-11;
; Matches 51; Conservative 1; Mismatches 0; Indels
;
; QY 1 GAKAPPAPAPQPGQPGQPGQPGQPGQPGQPGQPGQPEAPAPQPPAGG
;
; Db 1 GAKAPPAPAPAPQPGQPGQPGQPGQPGQPGQPGQPGQPEAPAPQPPAGG
;
;
; RESULT 10
; US-09-855-754B-14
;
; Sequence 14, Application US/09855754B
;
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE

```

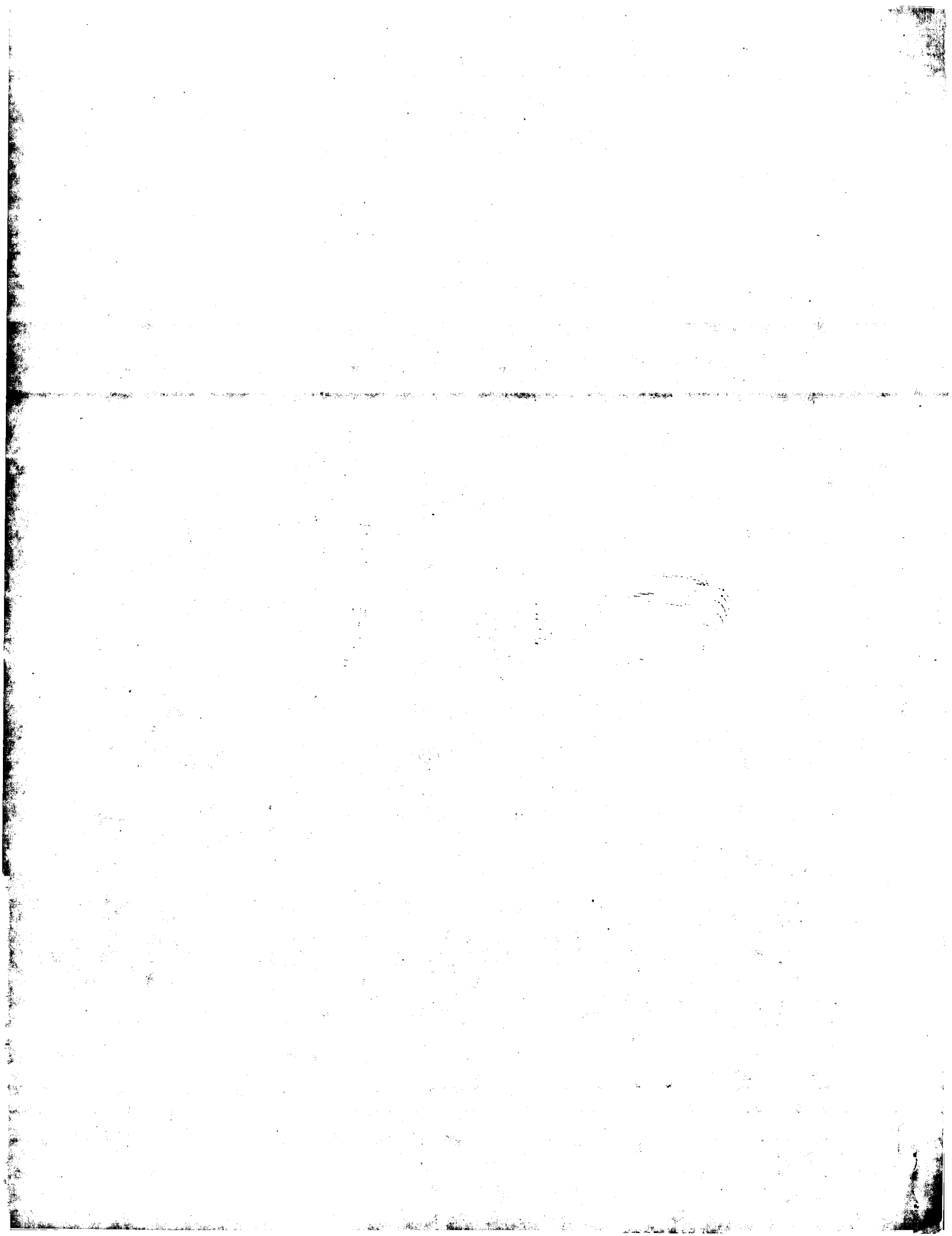




```

RESULT 8
US-10-227-353-4
; Sequence 4, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
; YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Milten, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

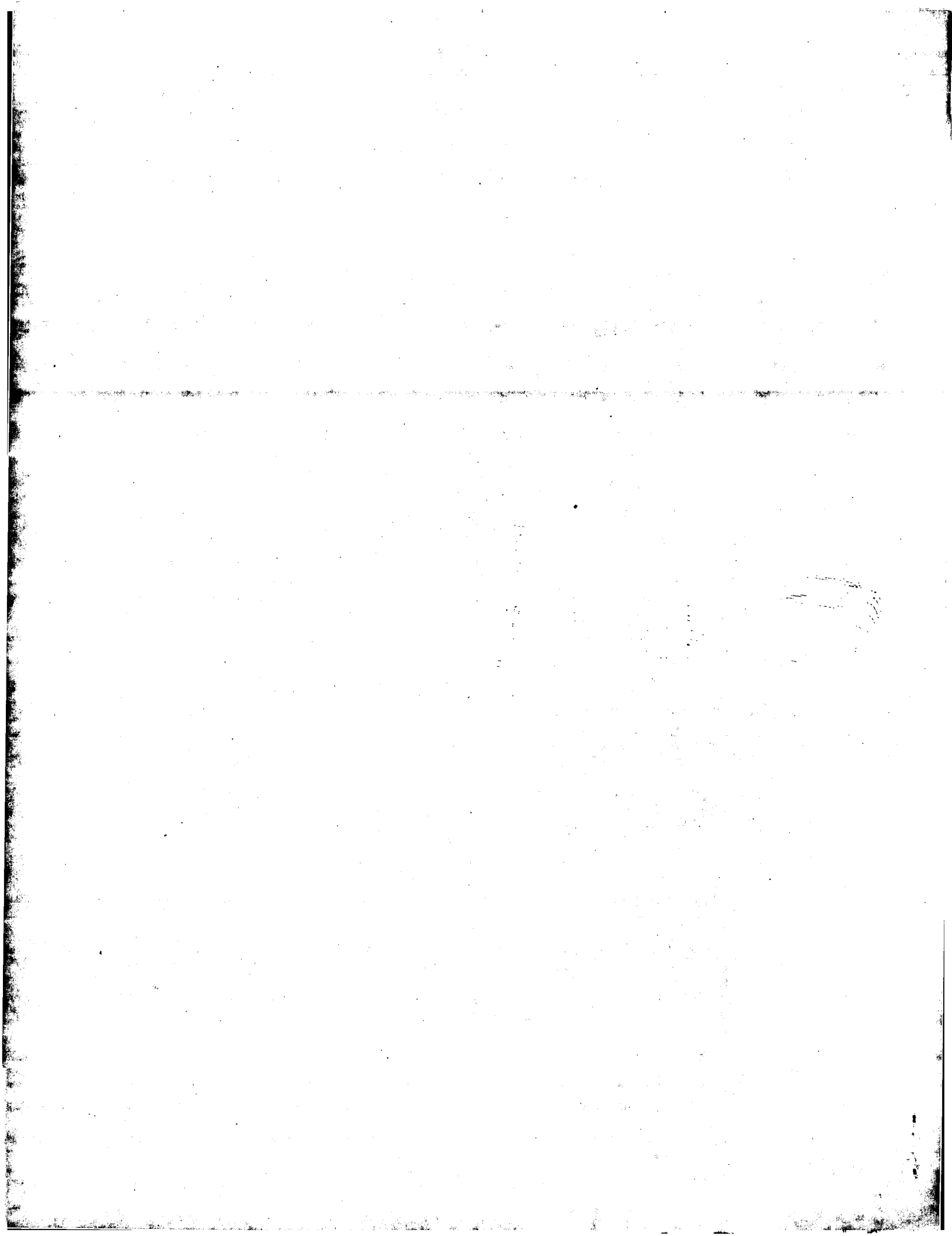
```

RESULT 10
US-09-855-754B-22
; Sequence 22, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAXX-EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE

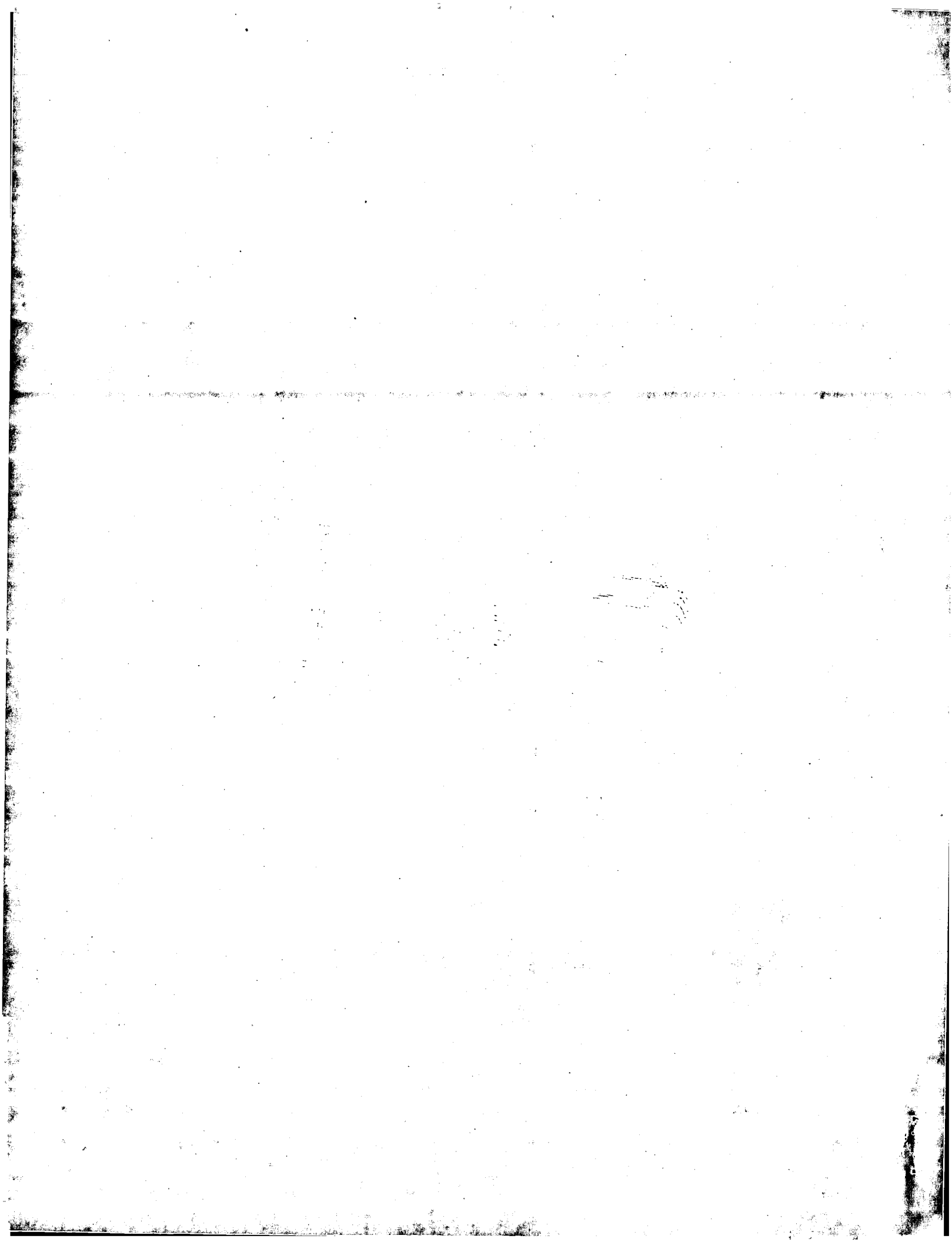


US-09-855-754-21
; Sequence 21, Application US/09855754

[illegible]



Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	272	100.0	48	5	US-09-855-754B-20		Sequence 20, Appl
2	261.5	96.1	49	5	US-09-855-754B-14		Sequence 14, Appl
3	260	95.6	52	5	US-09-855-754B-15		Sequence 15, Appl
4	260	95.6	911	5	US-09-855-754B-4		Sequence 4, Appl
5	260	95.6	911	6	US-10-227-353-4		Sequence 4, Appl
6	258	94.9	56	5	US-09-855-754B-18		Sequence 18, Appl
7	257	94.5	58	5	US-09-855-754B-19		Sequence 19, Appl
8	257	94.5	922	5	US-09-855-754B-6		Sequence 6, Appl
9	257	94.5	922	6	US-10-227-353-6		Sequence 6, Appl
10	249	91.5	52	5	US-09-855-754B-21		Sequence 21, Appl
11	248	91.2	54	5	US-09-855-754B-22		Sequence 22, Appl
12	244	89.7	52	5	US-09-855-754B-17		Sequence 17, Appl
13	240.5	88.4	59	5	US-09-855-754B-16		Sequence 16, Appl
14	223	82.0	42	5	US-09-855-754B-23		Sequence 23, Appl
15	223	82.0	768	6	US-10-283-123A-51012		Sequence 51012, A
16	223	82.0	910	5	US-09-855-754B-5		Sequence 5, Appl
17	223	82.0	910	6	US-10-227-353-2		Sequence 2, Appl
18	202.5	78.0	39	5	US-09-855-754B-24		Sequence 24, Appl
19	135.5	45.8	900	6	US-10-218-140-4170		Sequence 4170, Ap
20	135.5	49.8	1218	7	US-60-452-680-19541		Sequence 19541, A
21	135.5	49.8	1218	7	US-60-453-135-11684		Sequence 11684, A
22	135.5	49.8	1218	7	US-60-453-050-11684		Sequence 11684, A
23	129	47.4	523	6	US-10-017-161-1982		Sequence 1982, Ap
24	127	46.7	487	6	US-10-224-999A-3465		Sequence 3465, Ap
25	126	46.3	1076	5	US-09-724-676-84275		Sequence 84275, A
26	126	46.3	1076	5	US-09-724-676A-84275		Sequence 84275, A



sequence 37413, application 03/03/2133/

FC/CC000/00 NOTAPORT/Ddu /02 2000F00 /

[illegible]

; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 23
 ; LENGTH: 42
 ; TYPE: PRT
 ; ORGANISM: Bordetella bronchiseptica
 US-09-855-754-23

Query Match 72.8%; Score 217; DB 22; Length 42;
 Best Local Similarity 78.8%; Pred. No. 7.4e-08;
 Matches 41; Conservative 0; Mismatches 1; Indels 10; Gaps 1;

QY 1 GAKVPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 42
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: May 7, 2003, 17:22:08
 Job time : 141.727 secs

Result No.	Query				DB	ID	Description
	Score	Match	Length				
1	298	100.0	52	5	US-09-855-754B-21	Sequence 2	
2	287	96.3	54	5	US-09-855-754B-22	Sequence 2	
3	259	86.9	52	5	US-09-855-754B-15	Sequence 4	
4	259	86.9	911	6	US-09-855-754B-4	Sequence 4	
5	259	86.9	911	6	US-10-227-353-4	Sequence 4	
6	257	86.2	56	5	US-09-855-754B-18	Sequence 1	
7	256	85.9	58	5	US-09-855-754B-19	Sequence 1	
8	256	85.9	922	5	US-09-855-754B-6	Sequence 6	
9	256	85.9	922	6	US-10-227-353-6	Sequence 6	
10	249	83.6	48	5	US-09-855-754B-20	Sequence 2	
11	242.5	81.4	59	5	US-09-855-754B-16	Sequence 1	
12	238.5	80.0	49	5	US-09-855-754B-14	Sequence 1	
13	236	79.2	52	5	US-09-855-754B-17	Sequence 1	
14	217	72.8	42	5	US-09-855-754B-23	Sequence 2	
15	217	72.8	768	6	US-10-282-122A-51012	Sequence 5	
16	217	72.8	910	5	US-09-855-754B-5	Sequence 5	
17	217	72.8	910	6	US-10-227-353-2	Sequence 2	
18	196.5	65.9	39	5	US-09-855-754B-24	Sequence 2	
19	150	50.3	523	6	US-10-017-161-1982	Sequence 1	
20	148.5	49.8	900	6	US-10-218-140-4170	Sequence 4	
21	148.5	49.8	1218	7	US-60-452-680-19541	Sequence 1	
22	148.5	49.8	1218	7	US-60-453-135-11684	Sequence 1	
23	148.5	49.8	1218	7	US-60-453-050-11684	Sequence 1	
24	142	47.7	1444	5	US-09-949-016-9652	Sequence 9	
25	142	47.7	2004	5	US-09-949-016-6756	Sequence 6	
26	142	47.7	2004	7	US-60-423-586-80	Sequence 8	

Search completed: May 7, 2003, 17:27:51
Job time : 31.6508 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:47:16 ; Search time 136.794 Seconds
(without alignments)
254.511 Million cell updates/sec

Title: US-09-855-754B-22
Perfect score: 310
Sequence: 1 GAKVPPAPAPQGPQPQ.....QFQPEAPAPQAPAGRELSAA 54

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1:	/cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US098_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US099_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US100_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US101_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US102_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	310	100.0	54	22	US-09-855-754-22
2	287	92.6	52	22	US-09-855-754-21
3	265	85.5	58	22	US-09-855-754-19
4	265	85.5	922	21	US-09-791-537-50286
5	265	85.5	922	22	US-09-855-754-6
6	258	83.2	52	22	US-09-855-754-15

Sequence 22, Appl
Sequence 21, Appl
Sequence 19, Appl
Sequence 50286, A
Sequence 6, Appl
Sequence 15, Appl

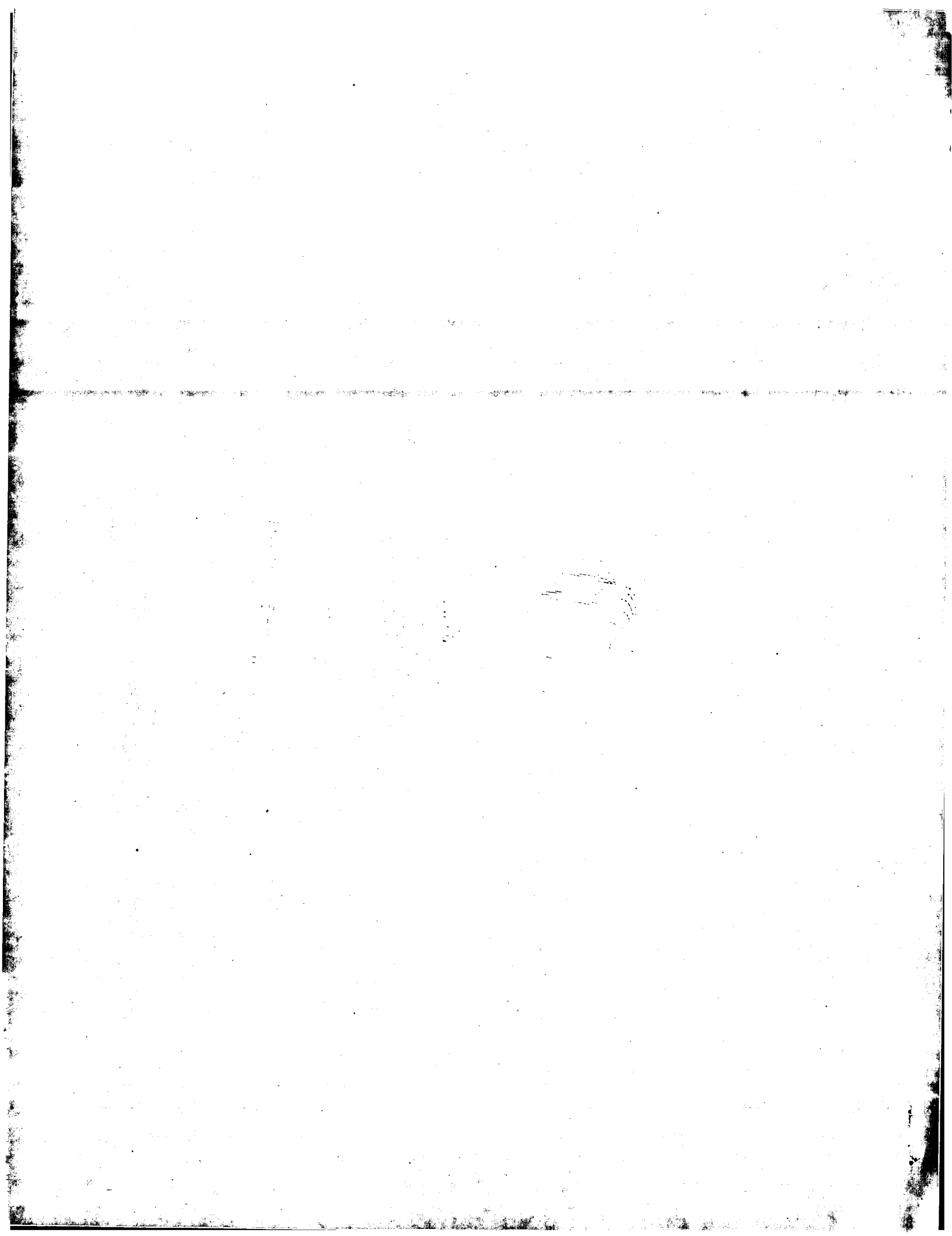
7	258	83.2	911	21	US-09-791-537-37413	Sequence 37413, A
8	258	83.2	911	21	US-09-791-537-85698	Sequence 85698, A
9	258	83.2	911	22	US-09-855-754-4	Sequence 4, Appl
10	256	82.6	56	22	US-09-855-754-18	Sequence 18, Appl
11	249.5	80.5	59	22	US-09-855-754-16	Sequence 16, Appl
12	248	80.0	48	22	US-09-855-754-20	Sequence 20, Appl
13	237.5	76.6	52	22	US-09-855-754-14	Sequence 14, Appl
14	235	75.8	52	22	US-09-855-754-17	Sequence 17, Appl
15	216	69.7	42	22	US-09-855-754-23	Sequence 23, Appl
16	216	69.7	138	4	US-08-030-424-2	Sequence 2, Appl
17	216	69.7	905	21	US-09-791-537-75841	Sequence 75841, A
18	216	69.7	905	21	US-09-791-537-76968	Sequence 76968, A
19	216	69.7	905	21	US-09-791-537-76981	Sequence 76981, A
20	216	69.7	910	21	US-09-791-537-21274	Sequence 21274, A
21	216	69.7	910	21	US-09-791-537-33443	Sequence 33443, A
22	216	69.7	910	21	US-09-791-537-98868	Sequence 98868, A
23	216	69.7	910	21	US-09-791-537-112007	Sequence 112007, A
24	216	69.7	910	22	US-09-855-754-5	Sequence 5, Appl
25	216	69.7	915	21	US-09-791-537-111629	Sequence 111629, A
26	197	63.5	39	6	US-08-247-544-2	Sequence 2, Appl
27	195.5	63.1	39	22	US-09-855-754-24	Sequence 24, Appl
28	195.5	63.1	907	21	US-09-791-537-78242	Sequence 78242, A
29	175	56.5	201	17	US-09-380-693A-49	Sequence 49, Appl
30	170.5	55.0	178	17	US-09-380-693A-51	Sequence 51, Appl
31	168	54.2	33	6	US-08-247-544-3	Sequence 3, Appl
32	163	52.6	406	17	US-09-380-693A-38	Sequence 38, Appl
33	159	51.3	174	17	US-09-380-693A-50	Sequence 50, Appl
34	158	51.0	31	6	US-08-247-544-8	Sequence 8, Appl
35	157.5	50.8	3164	27	US-60-412-956-16	Sequence 16, Appl
36	156	50.3	31	6	US-08-247-544-5	Sequence 5, Appl
37	155.5	50.2	825	27	US-60-389-987-2285	Sequence 2285, Ap
38	155.5	50.2	825	27	US-60-412-418-2285	Sequence 2285, A
39	155.5	50.2	1217	27	PCT-US01-08631-40235	Sequence 40235, A
40	155.5	50.2	1217	27	US-60-324-631-2916	Sequence 2916, Ap
41	155.5	50.2	1239	1	PCT-US01-08631-40236	Sequence 40236, A
42	155.5	50.2	1638	1	PCT-US02-30474-1318	Sequence 1318, Ap
43	155.5	50.2	1638	27	US-60-324-631-1323	Sequence 1323, Ap
44	153.5	49.5	171	17	US-09-380-693A-48	Sequence 48, Appl
45	152.5	49.2	279	18	US-09-417-507-22215	Sequence 22215, A

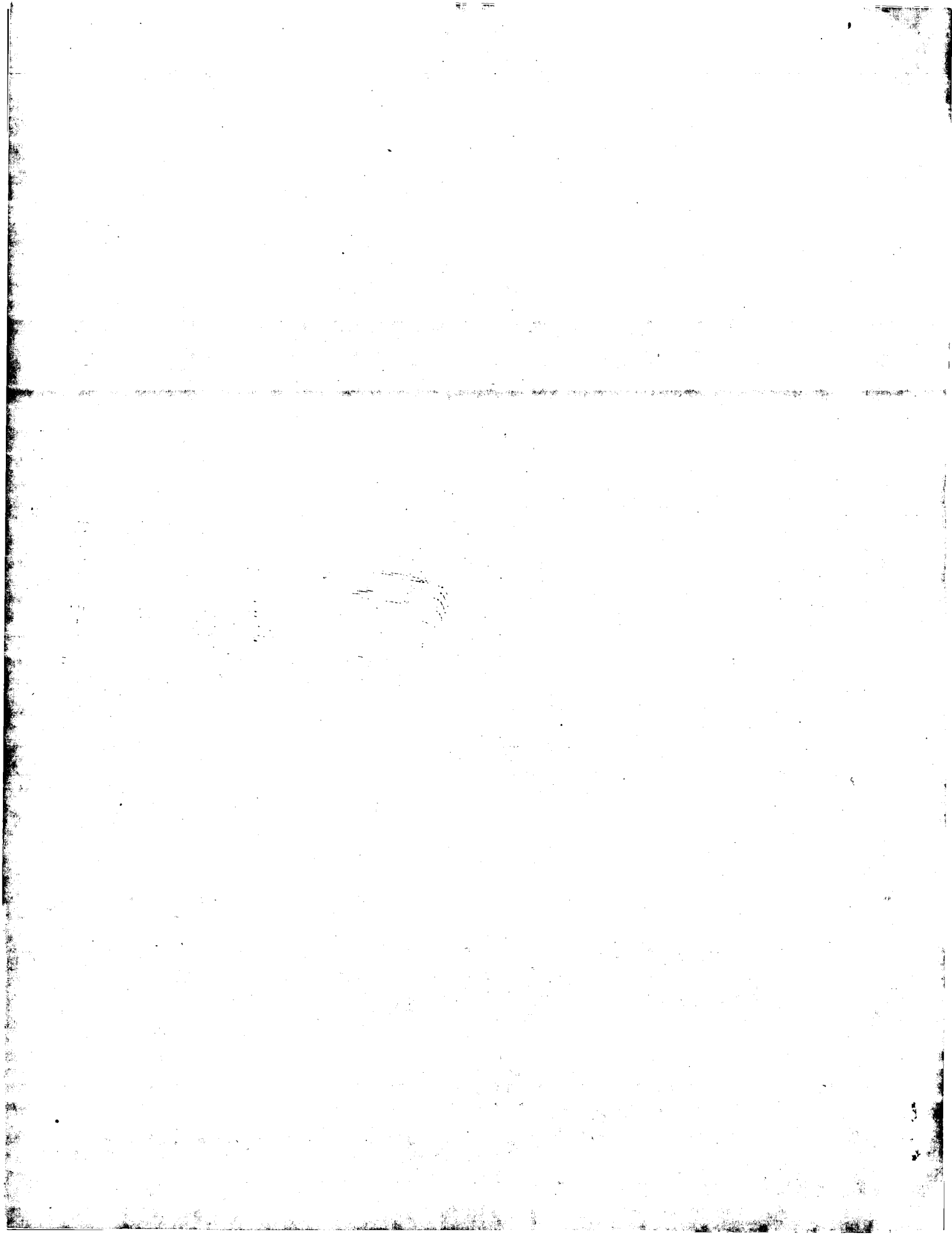
ALIGNMENTS

RESULT 1

US-09-855-754-22
; Sequence 22, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSAUX-UEDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLE PERTUSSIS,
; TITLE OF INVENTION: BORDETELLE PARAPERTUSSIS, AND BORDETELLE
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-22

Query Match 100.0%; Score 310; DB 22; Length 54;
Best Local Similarity 100.0%; Pred. No. 8.3e-14;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKVPPAPAPQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAGRELSAA 54





Result No.	Query #			ID	Description	
	Score	Match	Length			
1	297	100.0	52	5	US-09-855-754B-15	Sequence 15, Appl
2	297	100.0	911	5	US-09-855-754B-4	Sequence 4, Appl
3	297	100.0	911	6	US-10-227-353-4	Sequence 4, Appl
4	285	96.0	56	5	US-09-855-754B-18	Sequence 18, Appl
5	284	95.6	58	5	US-09-855-754B-19	Sequence 19, Appl
6	284	95.6	922	5	US-09-855-754B-6	Sequence 6, Appl
7	284	95.6	922	6	US-10-227-353-6	Sequence 6, Appl
8	271	91.2	52	5	US-09-855-754B-17	Sequence 17, Appl
9	267.5	90.1	59	5	US-09-855-754B-16	Sequence 16, Appl
10	266.5	89.7	49	5	US-09-855-754B-14	Sequence 14, Appl
11	260	87.5	48	5	US-09-855-754B-20	Sequence 20, Appl
12	259	87.2	52	5	US-09-855-754B-21	Sequence 21, Appl
13	258	86.9	54	5	US-09-855-754B-22	Sequence 22, Appl
14	212	71.4	42	5	US-09-855-754B-23	Sequence 23, Appl
15	212	71.4	768	6	US-10-289-122A-51012	Sequence 51012, A
16	212	71.4	910	5	US-09-855-754B-5	Sequence 5, Appl
17	212	71.4	910	6	US-10-227-353-2	Sequence 2, Appl
18	191.5	64.5	39	5	US-09-855-754B-24	Sequence 24, Appl
19	142.5	48.0	900	6	US-10-218-140-4170	Sequence 4170, Ap
20	142.5	48.0	1218	7	US-60-452-680-19541	Sequence 19541, A
21	142.5	48.0	1218	7	US-60-453-135-11684	Sequence 11684, A
22	142.5	48.0	1218	7	US-60-453-050-11684	Sequence 11684, A
23	140	47.1	523	6	US-10-017-161-1982	Sequence 1982, Ap
24	136	45.8	487	6	US-10-224-999A-3465	Sequence 3465, A
25	131	44.1	632	6	US-10-289-122A-51472	Sequence 51472, A
26	129	43.4	578	6	US-10-156-761-12544	Sequence 12544, A

[illegible]

RESULT 12
US-09-855-754B-21
; Sequence 21, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
CURRENT APPLICATION NUMBER: US/09/855,754B

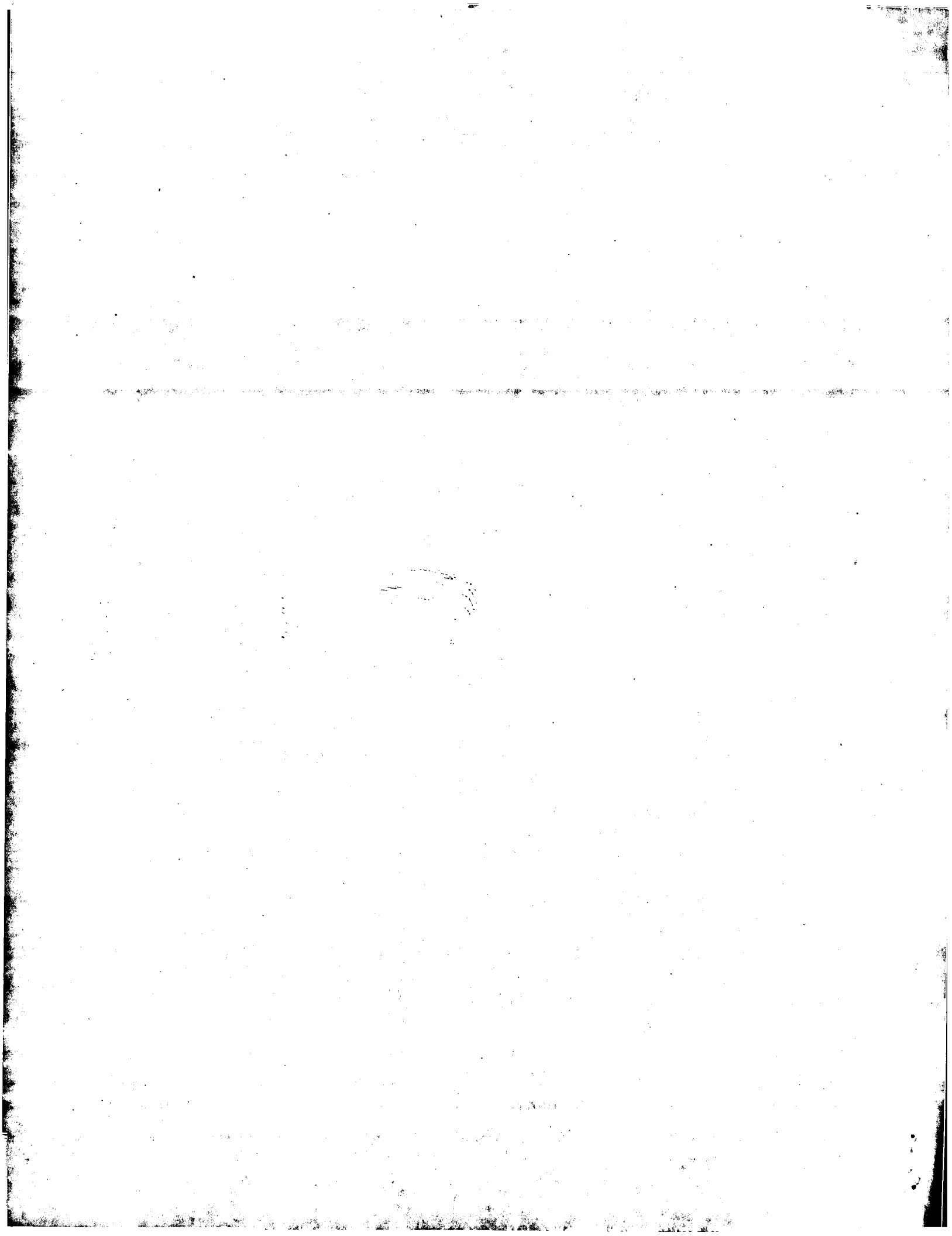
```

RESULT 14
US-09-855-754B-23
Sequence 23, Application US/09855754B
GENERAL INFORMATION:
APPLICANT: BOURSAAU-EUDE, CAROLINE
APPLICANT: GUISSO-MACLOUF, NICOLE
TITLE OF INVENTION: POLYPEPTIDES CONTAINING
TITLE OF INVENTION: REGIONS OF PERTURBED
TITLE OF INVENTION: BORDETELLA PARACITICULAE
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
FILE REFERENCE: 03495-0206-000000
CURRENT APPLICATION NUMBER: US/09/855754B
CURRENT FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/206,969
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 23
LENGTH: 42

```

0
0
0
0
-
4
0
-
0
.....
4
0

0
0
0



Result No.	Score	Query Match	Length	DB	ID	Description
1	297	100.0	52	22	US-09-855-754-15	Sequence 15, Appl
2	297	100.0	911	21	US-09-791-537-37413	Sequence 37413, A
3	297	100.0	911	21	US-09-791-537-85698	Sequence 85698, A
4	297	100.0	911	22	US-09-855-754-4	Sequence 4, Appl
5	285	96.0	56	22	US-09-855-754-18	Sequence 18, Appl
6	284	95.6	58	22	US-09-855-754-19	Sequence 19, Appl

```
Query Match      100.0% Score 297; DB 22; Length 52;
Best Local Similarity 100.0%; Pred. No. 5.3e-13;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


Qy **1 GAKAPPAPQPQGPPG-----PQPPQPQPQPQPQPQPQPQPAGRELSAA 52**

RESULT 11
US-09-855-754-14
; Sequence 14, Application US/098555754

```

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-21

Query Match      87.2%; Score 259; DB 22; Length 52;
Best Local Similarity 92.5%; Pred. No. 1.8e-10;
Matches 49; Conservative 0; Mismatches 2; Indels 2; Gaps 2

QY   1 GAKAPPAPKPAQPQGPPGPPGPPQP-PQRQPEAPAPQPPAGRELSAA 52
    ||| ||||||||| ||||||||| ||| ||||||||| ||| ||||||||| |||
Db    1 GAKVPPAPKPAQPQGPP-PQPPGPPGPPGPPQPPQPPQPEAPAPQPPAGRELSAA 52

RESULT 14
US-09-855-754-22
; Sequence 22, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPTERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHITSEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-22

Query Match      86.9%; Score 258; DB 22; Length 54;
Best Local Similarity 89.1%; Pred. No. 2.2e-10;
Matches 49; Conservative 0; Mismatches 2; Indels 4; Gaps 2

QY   1 GAKAPPAPKPAQPQGPPGPPGPPQP---PQRQPEAPAPQPPAGRELSAA 52
    ||| ||||||||| ||||||||| ||| ||||||||| ||| ||||||||| |||
Db    1 GAKVPPAPKPAQPQGPP-PQPPGPPGPPGPPQPPQPPQPEAPAPQPPAGRELSAA 54

RESULT 15
US-09-855-754-23
; Sequence 23, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUIISO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPTERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHITSEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
```



```
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-19

Query Match          94.8%; Score 263.5; DB 5; Length 58;
Best Local Similarity 84.5%; Pred. No. 1.2e-10;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPPAPKAPQPGPG-----PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 49
|||||
DB 1 GAKAPPAPKAPQPGPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 58
|||||

RESULT 7
US-09-855-754B-6
; Sequence 6, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-855-754B-6

Query Match          94.8%; Score 263.5; DB 5; Length 922;
Best Local Similarity 84.5%; Pred. No. 7.1e-10;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPPAPKAPQPGPG-----PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 49
|||||
DB 564 GAKAPPAPKAPQPGPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 621
|||||

RESULT 8
US-10-227-353-6
; Sequence 6, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; APPLICANT: ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Milten, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/227,353
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-227-353-6

Query Match          94.8%; Score 263.5; DB 6; Length 922;
Best Local Similarity 84.5%; Pred. No. 7.1e-10;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

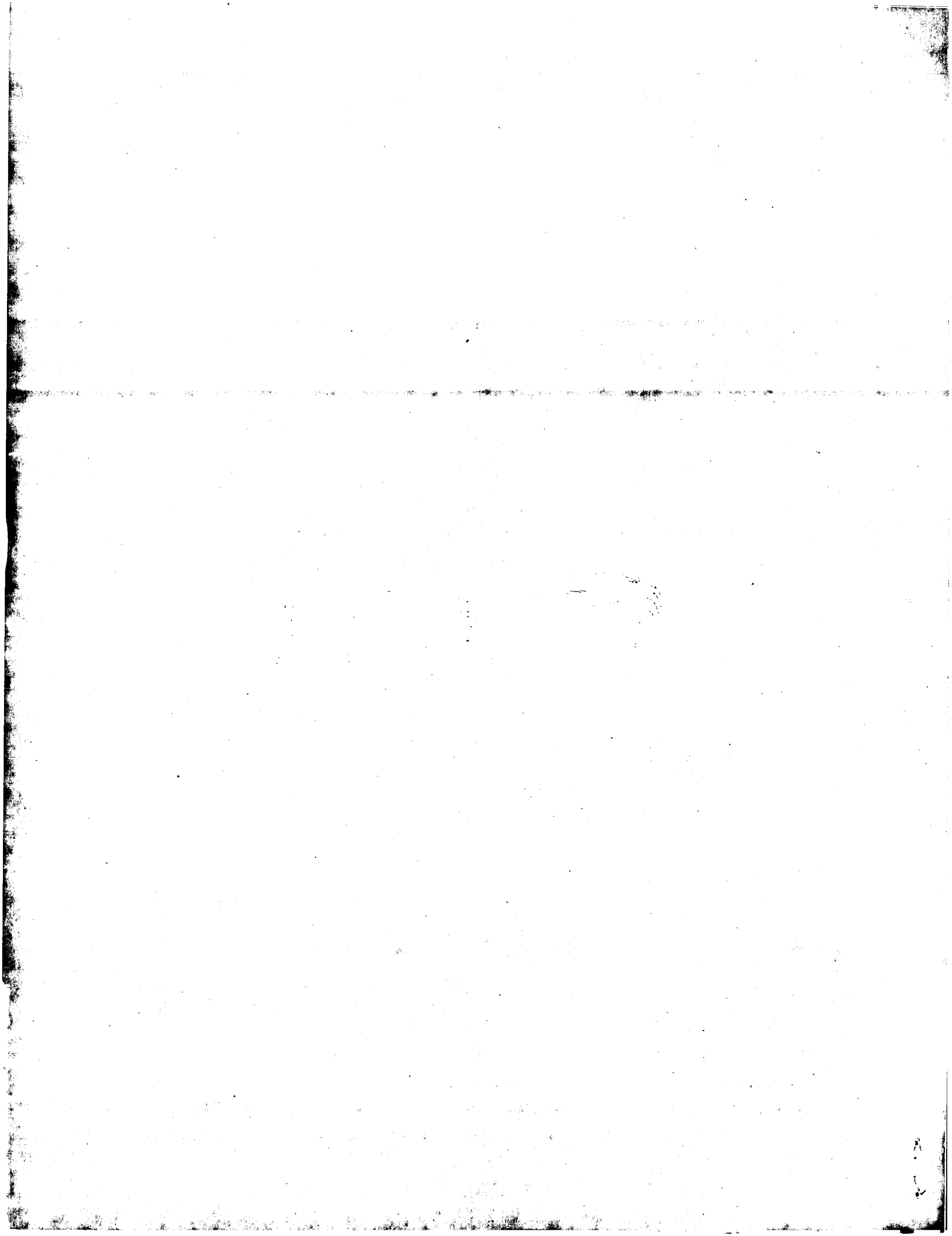
QY 1 GAKAPPAPKAPQPGPG-----PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 49
|||||
DB 564 GAKAPPAPKAPQPGPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 621
|||||
```

```
RESULT 9
US-09-855-754B-20
; Sequence 20, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-20

Query Match          94.1%; Score 261.5; DB 5; Length 48;
Best Local Similarity 98.0%; Pred. No. 1.5e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAKAPPAPKAPQPGPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 49
|||||
DB 1 GAKAPPAPKAPQPGQP-PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 48
|||||
```

```
RESULT 10
US-09-855-754B-17
; Sequence 17, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUIISO-MACLOUF, NICOLE
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:48:21 ; Search time 33.0085 Seconds
(without alignments)
289.347 Million cell updates/sec

Title: US-09-855-754B-9
Perfect score: 308
Sequence: 1 QRATIRGDAPAGGVPVGA.....PVLGHWYGVDSGSTVELAQ 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 785622 seqs, 170552221 residues

Total number of hits satisfying chosen parameters: 785622

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US05_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	308	100.0	56	5	US-09-855-754B-9
2	291	94.5	56	5	US-09-855-754B-12
3	281	91.2	56	5	US-09-855-754B-10
4	281	91.2	768	6	US-10-282-122A-51012
5	278.5	90.4	61	5	US-09-855-754B-11
6	274	89.0	910	5	US-09-855-754B-5
7	271	88.0	910	6	US-10-227-353-2
8	254.5	82.6	51	5	US-09-855-754B-13
9	228.5	74.2	51	5	US-09-855-754B-7
10	228.5	74.2	922	5	US-09-855-754B-6
11	228.5	74.2	922	6	US-10-227-353-6
12	212	68.8	46	5	US-09-855-754B-8
13	212	68.8	911	5	US-09-855-754B-4
14	212	68.8	911	6	US-10-227-353-4
15	112	36.4	171	1	PCT-US02-18256-45
16	112	36.4	2016	1	PCT-US02-18256-52
17	112	36.4	3420	1	PCT-US02-18256-51
18	111.5	36.2	63	1	PCT-US02-18256-46
19	96	31.2	1128	1	PCT-US02-18256-49
20	92.5	30.0	399	6	US-10-310-154-508
21	91.5	29.7	624	1	PCT-US02-18256-12
22	89.5	29.1	627	1	PCT-US03-00814-28
23	89.5	29.1	627	6	US-10-341-096-28
24	89.5	29.1	627	6	US-10-406-832-28
25	89.5	29.1	627	6	US-10-406-832-28
26	83	26.9	47	1	PCT-US02-18256-43

27	82	26.6	617	1	PCT-US03-01096-334	Sequence 334, App
28	82	26.6	759	5	US-09-724-676-72624	Sequence 72624, A
29	82	26.6	759	5	US-09-724-676A-72624	Sequence 72624, A
30	82	26.6	2234	5	US-09-724-676-72630	Sequence 72630, A
31	82	26.6	2234	5	US-09-724-676-72631	Sequence 72631, A
32	82	26.6	2234	5	US-09-724-676A-72630	Sequence 72630, A
33	82	26.6	2234	5	US-09-724-676A-72631	Sequence 72631, A
34	82	26.6	2270	5	US-09-724-676-72628	Sequence 72628, A
35	82	26.6	2270	5	US-09-724-676-72629	Sequence 72629, A
36	82	26.6	2270	5	US-09-724-676A-72628	Sequence 72628, A
37	82	26.6	2270	5	US-09-724-676A-72629	Sequence 72629, A
38	82	26.6	2910	7	US-60-453-135-13303	Sequence 13303, A
39	82	26.6	2910	7	US-60-453-050-13303	Sequence 13303, A
40	82	26.6	2911	6	US-10-126-052A-459	Sequence 459, App
41	82	26.6	2913	5	US-09-724-676-72627	Sequence 72627, A
42	82	26.6	2913	5	US-09-724-676A-72627	Sequence 72627, A
43	81	26.3	596	1	PCT-US02-18256-14	Sequence 14, Appl
44	80.5	26.1	417	7	US-60-453-135-8494	Sequence 8494, Ap
45	80.5	26.1	417	7	US-60-453-050-8494	Sequence 8494, Ap

ALIGNMENTS

RESULT 1

US-09-855-754B-9
; Sequence 9, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-9

Query Match 100.0%; Score 308; DB 5; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.7e-22;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRATIRGDAPAGGVPVGA
Db 1 QRATIRGDAPAGGVPVGA
|||||
|||||

RESULT 2

US-09-855-754B-12
; Sequence 12, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25

```
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-12

Query Match          94.5%; Score 291; DB 5; Length 56;
Best Local Similarity 94.6%; Pred. No. 1.9e-20;
Matches 53; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 56
|||||
Db 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 56
|||||

RESULT 3
US-09-855-754B-10
; Sequence 10, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-10

Query Match          91.2%; Score 281; DB 5; Length 56;
Best Local Similarity 92.9%; Pred. No. 1.6e-19;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 56
|||||
Db 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 56
|||||

RESULT 4
US-10-282-122A-51012
; Sequence 51012, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51012
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-10-282-122A-51012

Query Match          91.2%; Score 281; DB 6; Length 768;
Best Local Similarity 92.9%; Pred. No. 1.9e-18;
Matches 52; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 56
|||||
Db 112 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELAQ 167
|||||

RESULT 5
US-09-855-754B-11
; Sequence 11, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-11

Query Match          90.4%; Score 278.5; DB 5; Length 61;
Best Local Similarity 86.9%; Pred. No. 3e-19;
Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELA 55
|||||
Db 1 QRATIRRGDAPAGGVPGGVPGGFGDPGGFGPVLDDGWYGVDSGSTVELA 60
|||||

QY 56 Q 56
Db 61 Q 61

RESULT 6
```



```

Db 1 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFLDGGYGVDSSTVDLAQ 51
|||||
RESULT 10
US-09-855-754B-6
; Sequence 6, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-855-754B-6

Query Match 74.2%; Score 228.5; DB 5; Length 922;
Best Local Similarity 80.4%; Pred. No. 1.8e-13;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFLDGGYGVDSSTVDLAQ 56
|||||
Db 254 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFLDGGYGVDSSTVDLAQ 304
|||||

RESULT 11
US-10-227-353-6
; Sequence 6, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; APPLICANT: ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-227-353-6

Query Match 74.2%; Score 228.5; DB 6; Length 922;
Best Local Similarity 80.4%; Pred. No. 1.8e-13;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFLDGGYGVDSSTVDLAQ 56
|||||
Db 254 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFLDGGYGVDSSTVDLAQ 304
|||||

RESULT 12
US-09-855-754B-8
; Sequence 8, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-8

Query Match 68.8%; Score 212; DB 5; Length 46;
Best Local Similarity 75.0%; Pred. No. 3.8e-13;
Matches 42; Conservative 2; Mismatches 2; Indels 10; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFLDGGYGVDSSTVDLAQ 56
|||||
Db 1 QRATIRGDAPAGGAVPGGAV-----PGGFLDGGYGVDSSTVDLAQ 46
|||||

RESULT 13
US-09-855-754B-4
; Sequence 4, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-4

```

Query Match 68.8%; Score 212; DB 5; Length 911;
Best Local Similarity 75.0%; Pred. No. 6.3e-12;
Matches 42; Conservative 2; Mismatches 2; Indels 10; Gaps 1;

QY 1 QRATIRGDAPAGGVPGGVPGGFGFDPGGFGFGLDGGWYGVDSGSTVELAQ 56
|||||
Db 254 QRATIRGDAPAGGAVPGGAV-----PGGFGPLLDGGWYGVDSSTVDLAQ 299

RESULT 14

US-10-227-353-4

; Sequence 4, Application US/10227353

; GENERAL INFORMATION:

; APPLICANT: CLARE, JEFFREY J.

; ROMANOS, MICHAEL A.

; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.

; STREET: 2200 Clarendon Blvd., Suite 1400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/227,353

; FILING DATE: 26-Aug-2002

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,269C

; FILING DATE: 02-Jun-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Lebovitz, Richard M.

; REGISTRATION NUMBER: 37,067

; REFERENCE/DOCKET NUMBER: Popov-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 243-6333

; TELEFAX: (703) 243-6410

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 911 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-10-227-353-4

Query Match 68.8%; Score 212; DB 6; Length 911;
Best Local Similarity 75.0%; Pred. No. 6.3e-12;
Matches 42; Conservative 2; Mismatches 2; Indels 10; Gaps 1;

QY 1 QRATIRGDAPAGGVPGGVPGGFGFDPGGFGFGLDGGWYGVDSGSTVELAQ 56
|||||
Db 254 QRATIRGDAPAGGAVPGGAV-----PGGFGPLLDGGWYGVDSSTVDLAQ 299

RESULT 15

PCT-US02-18256-45

; Sequence 45, Application PC/TUS0218256

; GENERAL INFORMATION:

; APPLICANT: Don A. Roth

; APPLICANT: Randolph V. Lewis

; APPLICANT: The University of Wyoming

; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants

; FILE REFERENCE: UWYO 02-004

; CURRENT APPLICATION NUMBER: PCT/US02/18256

; CURRENT FILING DATE: 2002-06-06

; PRIOR APPLICATION NUMBER: 60/296,184

; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic spider silk protein repeat
PCT-US02-18256-45

Query Match 36.4%; Score 112; DB 1; Length 171;
Best Local Similarity 57.1%; Pred. No. 0.003;
Matches 24; Conservative 2; Mismatches 12; Indels 4; Gaps 2;

QY 8 GDAPAGGVPGGVPGGFGFDPGGFGFGLDGGWYGVDSG 49
|||||
Db 79 GSGPGGYG-PGGSGPGGYGPGGAGPGGSGP---GGYFGGSG 116

Search completed: May 7, 2003, 17:27:45

Job time : 34.0085 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:47:16 ; Search time 141.86 Seconds
(without alignments)
254.511 Million cell updates/sec

Title: US-09-855-754B-9
Perfect score: 308
Sequence: 1 QRATIRGDAPAGGVPVGGG.....PVLGDWYGVDSGSTVELAQ 56

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
 - 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
 - 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
 - 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
 - 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
 - 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
 - 27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	308	100.0	56	22	US-09-855-754-9
2	308	100.0	160	21	Sequence 9, Appli
3	308	100.0	161	21	Sequence 4640, Ap
4	291	94.5	56	22	Sequence 4644, Ap
5	291	94.5	910	21	Sequence 12, Appl
6	287	93.2	907	21	Sequence 112007, Sequence 78242, A

7	281	91.2	56	22	US-09-855-754-10
8	281	91.2	387	21	US-09-791-537-78243
9	281	91.2	910	21	US-09-791-537-21274
10	281	91.2	910	21	Sequence 21274, A
11	278.5	90.4	61	22	Sequence 33443, A
12	278.5	90.4	915	21	Sequence 11, Appl
13	274	89.0	910	22	Sequence 111629, Sequence 5, Appli
14	271	88.0	539	21	Sequence 31601, A
15	271	88.0	910	21	Sequence 98868, A
16	255.5	83.0	905	21	Sequence 76981, A
17	254.5	82.6	51	22	Sequence 13, Appl
18	254.5	82.6	905	21	Sequence 75841, A
19	234.5	82.6	905	21	Sequence 76968, A
20	238.5	74.2	51	22	Sequence 7, Appli
21	238.5	74.2	158	21	Sequence 4609, Ap
22	238.5	74.2	159	21	Sequence 4603, Ap
23	238.5	74.2	160	21	Sequence 4519, Ap
24	238.5	74.2	922	21	Sequence 50286, A
25	238.5	74.2	922	22	Sequence 6, Appli
26	212	68.8	46	22	Sequence 8, Appli
27	212	68.8	147	21	Sequence 4547, Ap
28	212	68.8	150	21	Sequence 4540, Ap
29	212	68.8	150	21	Sequence 4562, Ap
30	212	68.8	154	21	Sequence 4573, Ap
31	212	68.8	155	21	Sequence 4568, Ap
32	212	68.8	911	21	Sequence 37413, A
33	212	68.8	911	21	Sequence 85698, A
34	212	68.8	911	22	Sequence 4, Appli
35	120.5	39.1	907	14	Sequence 4, Appli
36	116	37.7	745	14	Sequence 28, Appl
37	116	37.7	745	14	Sequence 2, Appli
38	106	34.4	783	21	Sequence 49010, A
39	104	33.8	445	20	Sequence 15996, A
40	104	33.8	445	20	Sequence 795, App
41	104	33.8	445	27	Sequence 16037, A
42	104	33.8	445	27	Sequence 1275, Ap
43	104	33.8	445	27	Sequence 13133, A
44	104	33.8	445	27	Sequence 16040, A
45	104	33.8	445	27	Sequence 12698, A

ALIGNMENTS

RESULT 1

- US-09-855-754-9
- Sequence 9, Application US/09855754
- GENERAL INFORMATION:
- APPLICANT: BOURSAX-UEDE, CAROLINE
- APPLICANT: GUISSO-MACLOUF, NICOLE
- TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS, AND BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN IMMUNOGENIC COMPOSITIONS
- FILE REFERENCE: 03495-0206-00000
- CURRENT APPLICATION NUMBER: US/09/855.754
- CURRENT FILING DATE: 2001-09-10
- PRIOR APPLICATION NUMBER: 60/206,969
- PRIOR FILING DATE: 2000-05-25
- NUMBER OF SEQ ID NOS: 24
- SOFTWARE: PatentIn Ver. 2.1
- SEQ ID NO 9
- LENGTH: 56
- TYPE: PRT
- ORGANISM: Bordetella bronchiseptica
- US-09-855-754-9

Query Match 100.0%; Score 308; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.1e-24;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 QRATIRGDAPAGGVPVGGFGPGGFPVLGDWYGVDSGSTVELAQ 56

LENGTH: 61
 TYPE: PRT
 ORGANISM: Bordetella bronchiseptica
 US-09-855-754-11

Query Match 90.4%; Score 278.5; DB 22; Length 61;
 Best Local Similarity 86.9%; Pred. No. 2.6e-21;
 Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGVPAGV-----PGGFDGGGPGGPGVLDGWYGVDSGSTVELA 55
 DB 1 QRATIRGDAPAGGAVPGGAVPGGPGGPGGPGGPGVLDGWYGVDSGSTVELA 60

QY 56 Q 56
 DB 61 Q 61

RESULT 12
 US-09-791-537-111629
 ; Sequence 111629, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 111629
 ; LENGTH: 915
 ; TYPE: PRT
 ; ORGANISM: Bordetella pertussis
 US-09-791-537-111629

Query Match 90.4%; Score 278.5; DB 21; Length 915;
 Best Local Similarity 86.9%; Pred. No. 4.6e-20;
 Matches 53; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGVPAGV-----PGGFDGGGPGGPGVLDGWYGVDSGSTVELA 55
 DB 254 QRATIRGDAPAGGAVPGGAVPGGPGGPGGPGGPGVLDGWYGVDSGSTVELA 313

QY 56 Q 56
 DB 314 Q 314

RESULT 13
 US-09-855-754-5
 ; Sequence 5, Application US/09855754
 ; GENERAL INFORMATION:
 ; APPLICANT: BOURSNAUX-EUDE, CAROLINE
 ; APPLICANT: GUISSO-MACLOUF, NICOLE
 ; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
 ; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
 ; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
 ; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
 ; FILE REFERENCE: 03495-0206-00000
 ; CURRENT APPLICATION NUMBER: US/09/855,754
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/206,969
 ; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 910
 ; TYPE: PRT
 ; ORGANISM: Bordetella pertussis

US-09-855-754-5

Query Match 89.0%; Score 274; DB 22; Length 910;
 Best Local Similarity 91.1%; Pred. No. 1.3e-19;
 Matches 51; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGVPAGGAVPGGFDGGGPGGPGVLDGWYGVDSGSTVELAQ 56
 DB 254 QRATIRGDAPAGGAVPGGAVPGGPGGPGGPGVLDGWYGVDSGSTVELAQ 309

RESULT 14
 US-09-791-537-31601
 ; Sequence 31601, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 31601
 ; LENGTH: 539
 ; TYPE: PRT
 ; ORGANISM: pdb 1DABA
 US-09-791-537-31601

Query Match 88.0%; Score 271; DB 21; Length 539;
 Best Local Similarity 91.1%; Pred. No. 1.6e-19;
 Matches 51; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGVPAGGAVPGGFDGGGPGGPGVLDGWYGVDSGSTVELAQ 56
 DB 220 QRATIRGDALAGGAVPGGAVPGGPGGPGGPGVLDGWYGVDSGSTVELAQ 275

RESULT 15
 US-09-791-537-98868
 ; Sequence 98868, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 98868
 ; LENGTH: 910
 ; TYPE: PRT
 ; ORGANISM: Bordetella pertussis
 US-09-791-537-98868

Query Match 88.0%; Score 271; DB 21; Length 910;
 Best Local Similarity 91.1%; Pred. No. 2.7e-19;
 Matches 51; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGVPAGGAVPGGFDGGGPGGPGVLDGWYGVDSGSTVELAQ 56
 DB 254 QRATIRGDALAGGAVPGGAVPGGPGGPGGPGVLDGWYGVDSGSTVELAQ 309

Search completed: May 7, 2003, 17:21:52
 Job time : 142.86 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:48:21 ; Search time 30.0613 Seconds
(without alignments)
289.347 Million cell updates/sec

Title: US-09-855-754B-7
Perfect score: 272
Sequence: 1 QRATIRGDAPAGGAVPGGA.....PLLDGHWGVDSSTVDLAQ 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 785622 seqs, 170552221 residues

Total number of hits satisfying chosen parameters: 785622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US05_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272	100.0	51	5	US-09-855-754B-7
2	272	100.0	922	5	US-09-855-754B-6
3	272	100.0	922	6	US-10-227-353-6
4	248.5	91.4	910	5	US-09-855-754B-5
5	241.5	88.8	56	5	US-09-855-754B-10
6	241.5	88.8	788	6	US-10-282-122A-51012
7	241	88.6	51	5	US-09-855-754B-13
8	232.5	85.5	46	5	US-09-855-754B-8
9	232.5	85.5	911	6	US-09-855-754B-4
10	232.5	85.5	911	6	US-10-227-353-4
11	231.5	85.1	910	6	US-10-227-353-2
12	228.5	84.0	56	5	US-09-855-754B-9
13	228.5	84.0	56	5	US-09-855-754B-12
14	226	83.1	61	5	US-09-855-754B-11
15	82	30.1	171	1	PCT-US02-18256-45
16	82	30.1	3420	1	PCT-US02-18256-51
17	79	29.0	399	6	US-10-310-154-508
18	78.5	28.9	2016	1	PCT-US02-18256-52
19	77	28.3	63	1	PCT-US02-18256-46
20	73.5	27.0	542	6	US-10-366-683-26256
21	73.5	27.0	542	6	US-10-419-128-26256
22	70	25.7	613	6	US-10-282-122A-61755
23	70	25.7	1029	6	US-10-156-761-11059
24	69.5	25.6	162	1	PCT-US02-32727-24021
25	69.5	25.6	162	5	US-09-978-825-24021
26	69.5	25.6	162	6	US-10-057-498-24021

27	69.5	25.6	177	6	US-10-366-683-27635
28	69.5	25.6	177	6	US-10-419-128-27635
29	69.5	25.6	315	6	US-10-366-683-31381
30	69.5	25.6	315	6	US-10-419-128-31381
31	68	25.0	627	1	PCT-US03-00814-28
32	68	25.0	627	6	US-10-341-096-28
33	68	25.0	627	6	US-10-341-097-28
34	68	25.0	627	6	US-10-406-832-28
35	67.5	24.8	428	6	US-10-156-761-11698
36	66.5	24.4	47	6	US-10-276-774-2507
37	66.5	24.4	179	6	US-10-017-161-1200
38	66.5	24.4	197	6	US-10-366-683-30130
39	66.5	24.4	197	6	US-10-419-128-30130
40	66.5	24.4	233	6	US-10-156-761-12230
41	66.5	24.4	558	6	US-10-366-683-19536
42	66	24.3	412	6	US-10-419-128-19536
43	66	24.3	412	6	US-10-282-122A-47675
44	66	24.3	1001	6	US-10-156-761-10088
45	66	24.3	1046	6	US-10-156-761-10088

ALIGNMENTS

RESULT 1

US-09-855-754B-7
; Sequence 7, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; FILE REFERENCE: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-7

Query Match 100.0%; Score 272; DB 5; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.1e-20;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGGLLDGHWGVDSSTVDLAQ 51
|||||
Db 1 QRATIRGDAPAGGAVPGGAVPGGGLLDGHWGVDSSTVDLAQ 51

RESULT 2

US-09-855-754B-6
; Sequence 6, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; FILE REFERENCE: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25

```

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-855-754B-6

Query Match      100.0%; Score 272; DB 5; Length 922;
Best Local Similarity 100.0%; Pred. No. 3.7e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWDVSDSTVDLAQ 51
|||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWDVSDSTVDLAQ 304
|||||

RESULT 3
US-10-227-353-6
; Sequence 6, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-227-353-6

Query Match      100.0%; Score 272; DB 6; Length 922;
Best Local Similarity 100.0%; Pred. No. 3.7e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWDVSDSTVDLAQ 51
|||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWDVSDSTVDLAQ 304
|||||

RESULT 4
US-09-855-754B-5
; Sequence 5, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-09-855-754B-5

Query Match      91.4%; Score 248.5; DB 5; Length 910;
Best Local Similarity 85.7%; Pred. No. 8.2e-17;
Matches 48; Conservative 3; Mismatches 0; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWDVSDSTVDLAQ 51
|||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWDVSDSSVELAQ 309
|||||

RESULT 5
US-09-855-754B-10
; Sequence 10, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-10

Query Match      88.8%; Score 241.5; DB 5; Length 56;
Best Local Similarity 83.9%; Pred. No. 2.6e-17;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWDVSDSTVDLAQ 51
|||||
Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGFGPLLDGWDVSDSSVELAQ 56
|||||

RESULT 6
US-10-282-122A-51012
; Sequence 51012, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 51012
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-10-282-122A-51012

Query Match      88.8%; Score 241.5; DB 6; Length 768;
Best Local Similarity 83.9%; Pred. No. 3.5e-16;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 51
    |||||||
Db 112 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 167
    |||||||

RESULT 7
US-09-855-754B-13
; Sequence 13, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-13

Query Match      88.6%; Score 241; DB 5; Length 51;
Best Local Similarity 88.2%; Pred. No. 2.6e-17;
```

```
Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 51
    |||||||
Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 51
    |||||||

RESULT 8
US-09-855-754B-8
; Sequence 8, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-8

Query Match      85.5%; Score 232.5; DB 5; Length 46;
Best Local Similarity 90.2%; Pred. No. 1.7e-16;
Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 51
    |||||||
Db 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 46
    |||||||

RESULT 9
US-09-855-754B-4
; Sequence 4, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-4

Query Match      85.5%; Score 232.5; DB 5; Length 911;
Best Local Similarity 90.2%; Pred. No. 3.3e-15;
Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 51
    |||||||
Db 254 QRATIRGDAPAGGAVPGGAVPGGAVPGGPGPLDGGYGVDSSTVDLAQ 299
    |||||||
```

```
RESULT 10
US-10-227-353-4
; Sequence 4, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-227-353-4
Query Match 85.5%; Score 232.5; DB 6; Length 911;
Best Local Similarity 90.2%; Pred. No. 3.3e-15;
Matches 46; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFGPLLDGWGVDVSDSTDVLAQ 51
|||||
Db 254 QRATIRGDAPA-----GGAVPGGAVPGGFGPLLDGWGVDVSDSTDVLAQ 299
|||||

RESULT 11
US-10-227-353-2
; Sequence 2, Application US/10227353
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,353
; FILING DATE: 26-Aug-2002
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 910 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-227-353-2
Query Match 85.1%; Score 231.5; DB 6; Length 910;
Best Local Similarity 82.1%; Pred. No. 4.1e-15;
Matches 46; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAVPGGAV-----PGGFGPLLDGWGVDVSDSTDVLAQ 51
|||||
Db 254 QRATIRGDALAGGAVPGGAVPGGAVPGGFGFGPLLDGWGVDVSGSSVELAQ 309
|||||

RESULT 12
US-09-855-754B-9
; Sequence 9, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSAUX-UEDE, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-9
Query Match 84.0%; Score 228.5; DB 5; Length 56;
Best Local Similarity 80.4%; Pred. No. 5.2e-16;
Matches 45; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFGPLLDGWGVDVSDSTDVLAQ 51
|||||
Db 1 QRATIRGDAPAGGAVPGGAVPGGFGFGPLLDGWGVDVSGSTVELAQ 56
|||||

RESULT 13
US-09-855-754B-12
; Sequence 12, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSAUX-UEDE, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
```

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-12

Query Match 84.0%; Score 228.5; DB 5; Length 56;
Best Local Similarity 80.4%; Pred. No. 5.2e-16;
Matches 45; Conservative 3; Mismatches 3; Indels 5; Gaps 1;
QY 1 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFGPLLDGWYGVDSSTVDLAQ 51
Db 1 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGVLDGWYGVDSSTVELAQ 56

RESULT 14
US-09-855-754B-11
; Sequence 11, Application US/09855754B
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-EUDE, CAROLINE
; APPLICANT: GUIO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754B
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754B-11

Query Match 83.1%; Score 226; DB 5; Length 61;
Best Local Similarity 73.8%; Pred. No. 1e-15;
Matches 45; Conservative 3; Mismatches 3; Indels 10; Gaps 1;
QY 1 QRATIRGDAPAGGAVPGGAV-----PGGAVPGGFGPLLDGWYGVDSSTVDLA 50
Db 1 QRATIRGDAPAGGAVPGGAVPGGFGPGGFGVLDGWYGVDSSTVELA 60

QY 51 Q 51
Db 61 Q 61

RESULT 15
PCT-US02-18256-45
; Sequence 45, Application PC/TUS0218256
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; TITLE OF INVENTION: The University of Wyoming
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: UWYO 02-004
; CURRENT APPLICATION NUMBER: PCT/US02/18256
; CURRENT FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184

; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic spider silk protein repeat
PCT-US02-18256-45

Query Match 30.1%; Score 82; DB 1; Length 171;
Best Local Similarity 60.0%; Pred. No. 0.74;
Matches 15; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 8 GDAPAGGAVPGGAVPGGAVPGGFGP 32
Db 88 GGSPPGGYGPAGGAGPGSGPGGYGP 112

Search completed: May 7, 2003, 17:27:44
Job time: 32.0613 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:47:16 ; Search time 129.194 Seconds
(without alignments)
254.511 Million cell updates/sec

Title: US-09-855-754B-7
Perfect score: 272
Sequence: 1 QRATIRGDPAGGAVPGCA.....PLLDGWYGVDSSTVDLAQ 51

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pcp.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pcp.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pcp.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pcp.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pcp.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pcp.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pcp.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pcp.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pcp.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pcp.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pcp.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pcp.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pcp.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pcp.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pcp.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pcp.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pcp.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pcp.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pcp.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pcp.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pcp.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pcp.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pcp.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pcp.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pcp.*
- 26: /cgn2_6/ptodata/1/paa/US102_COMB.pcp.*
- 27: /cgn2_6/ptodata/1/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	100.0	51	22	US-09-855-754-7
2	272	100.0	158	21	US-09-791-537-4609
3	272	100.0	159	21	US-09-791-537-4603
4	272	100.0	160	21	US-09-791-537-4519
5	272	100.0	922	21	US-09-791-537-50286
6	272	100.0	922	22	US-09-855-754-6

7	248.5	91.4	910	22	US-09-855-754-5	Sequence 5, Appl1
8	241.5	88.8	56	22	US-09-855-754-10	Sequence 10, Appl1
9	241.5	88.8	387	21	US-09-791-537-78243	Sequence 78243, A
10	241.5	88.8	910	21	US-09-791-537-21274	Sequence 21274, A
11	241.5	88.8	910	21	US-09-791-537-33443	Sequence 33443, A
12	241.5	88.8	51	22	US-09-855-754-13	Sequence 13, Appl1
13	241.5	88.6	905	21	US-09-791-537-73841	Sequence 73841, A
14	241.5	88.6	905	21	US-09-791-537-76968	Sequence 76968, A
15	237.5	87.3	907	21	US-09-791-537-78242	Sequence 78242, A
16	232.5	85.5	46	22	US-09-855-754-8	Sequence 8, Appl1
17	232.5	85.5	147	21	US-09-791-537-4547	Sequence 4547, Ap
18	232.5	85.5	150	21	US-09-791-537-4540	Sequence 4540, Ap
19	232.5	85.5	150	21	US-09-791-537-4562	Sequence 4562, Ap
20	232.5	85.5	154	21	US-09-791-537-4573	Sequence 4573, Ap
21	232.5	85.5	155	21	US-09-791-537-4568	Sequence 4568, Ap
22	232.5	85.5	911	21	US-09-791-537-37413	Sequence 37413, A
23	232.5	85.5	911	21	US-09-791-537-85698	Sequence 85698, A
24	232.5	85.5	911	22	US-09-855-754-4	Sequence 4, Appl1
25	231.5	85.1	539	21	US-09-791-537-31601	Sequence 31601, A
26	231.5	85.1	910	21	US-09-791-537-98868	Sequence 98868, A
27	228.5	84.0	56	22	US-09-855-754-9	Sequence 9, Appl1
28	228.5	84.0	56	22	US-09-855-754-12	Sequence 12, Appl1
29	228.5	84.0	160	21	US-09-791-537-4640	Sequence 4640, Ap
30	228.5	84.0	161	21	US-09-791-537-4644	Sequence 4644, Ap
31	228.5	84.0	910	21	US-09-791-537-112007	Sequence 112007, A
32	228.5	84.0	905	21	US-09-791-537-76981	Sequence 76981, A
33	226	83.1	61	22	US-09-855-754-11	Sequence 11, Appl1
34	226	83.1	915	21	US-09-791-537-111629	Sequence 111629, A
35	94.5	34.7	745	14	US-09-010-928A-28	Sequence 28, Appl1
36	94.5	34.7	870	14	US-09-010-928A-2	Sequence 2, Appl1
37	83.5	30.7	907	14	US-09-010-928A-4	Sequence 4, Appl1
38	83.5	30.7	198	18	US-09-417-507-27641	Sequence 27641, A
39	83	30.5	783	21	US-09-791-537-49010	Sequence 49010, A
40	81	29.8	350	19	US-09-595-328C-2166	Sequence 2166, Ap
41	81	29.8	398	20	US-09-614-150-16389	Sequence 16389, A
42	81	29.8	398	21	US-09-791-537-55210	Sequence 55210, A
43	81	29.8	398	27	US-60-167-217-16419	Sequence 16419, A
44	81	29.8	398	27	US-60-173-464-13448	Sequence 13448, A
45	81	29.8	398	27	US-60-191-637-16433	Sequence 16433, A

ALIGNMENTS

RESULT 1

US-09-855-754-7
; Sequence 7, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, NICOLE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-7

Query Match 100.0%; Score 272; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.6e-22;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRATIRGDPAGGAVPGGAVPGGFLDGDWYGVDSSTVDLAQ 51

Db 1 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 51
|||||

RESULT 2

US-09-791-537-4609
; Sequence 4609, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4609
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-791-537-4609

Query Match 100.0%; Score 272; DB 21; Length 158;
Best Local Similarity 100.0%; Pred. No. 5.6e-22;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 51
|||||

Db 54 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 104
|||||

RESULT 3

US-09-791-537-4603
; Sequence 4603, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4603
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-791-537-4603

Query Match 100.0%; Score 272; DB 21; Length 159;
Best Local Similarity 100.0%; Pred. No. 5.6e-22;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 51
|||||

Db 53 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 103
|||||

RESULT 4

US-09-791-537-4519
; Sequence 4519, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4519
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-791-537-4519

Query Match 100.0%; Score 272; DB 21; Length 160;
Best Local Similarity 100.0%; Pred. No. 5.6e-22;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 51
|||||

Db 54 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 104
|||||

RESULT 5

US-09-791-537-50286
; Sequence 50286, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50286
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-791-537-50286

Query Match 100.0%; Score 272; DB 21; Length 922;
Best Local Similarity 100.0%; Pred. No. 4.1e-21;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 51
|||||

Db 254 QRATIRGDAPAGGAVPGGAVPGGFPLLDGGWYGVDSSTVDLAQ 304
|||||

RESULT 6

US-09-855-754-6
; Sequence 6, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bordetella parapertussis
US-09-855-754-6

Query Match 100.0%; Score 272; DB 22; Length 922;
Best Local Similarity 100.0%; Pred. No. 4.1e-21;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 51
|||||
Db 254 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 304

RESULT 7

US-09-855-754-5
; Sequence 5, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUIISO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-09-855-754-5

Query Match 91.4%; Score 248.5; DB 22; Length 910;
Best Local Similarity 85.7%; Pred. No. 1.6e-18;
Matches 48; Conservative 3; Mismatches 0; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 51
|||||
Db 254 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 309

RESULT 8

US-09-855-754-10
; Sequence 10, Application US/09855754
; GENERAL INFORMATION:
; APPLICANT: BOURSAX-UEDE, CAROLINE
; APPLICANT: GUIISO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETTELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETTELLA PARAPERTUSSIS, AND BORDETTELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-10

Query Match 88.8%; Score 241.5; DB 22; Length 56;
Best Local Similarity 83.9%; Pred. No. 4.2e-19;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 51
|||||
Db 1 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 56

RESULT 9

US-09-791-537-78243
; Sequence 78243, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 78243
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-09-791-537-78243

Query Match 88.8%; Score 241.5; DB 21; Length 387;
Best Local Similarity 83.9%; Pred. No. 3.7e-18;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 51
|||||
Db 254 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 309

RESULT 10

US-09-791-537-21274
; Sequence 21274, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21274
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-09-791-537-21274

Query Match 88.8%; Score 241.5; DB 21; Length 910;
Best Local Similarity 83.9%; Pred. No. 9.7e-18;
Matches 47; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

QY 1 QRATIRRGDAPAGGAVPGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 51
|||||
Db 254 QRATIRRGDAPAGGAVPGGAVPGGFPGLDGGYGVDSSTVDLAQ 309

RESULT 11

US-09-791-537-33443
; Sequence 33443, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33443

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.24361 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-22

Perfect score: 310

Sequence: 1 GAKVPPAPAPQGPQPQ.....QPOPEAPQPPAGRELSAA 54

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	265	85.5	922	1	PERT_BORPA
2	258	83.2	911	1	PERT_BORBR
3	216	69.7	910	1	PERT_BORPE
4	157.5	50.8	3164	1	TEGU_HSV11
5	153	49.4	485	1	SSGP_VOLCA
6	151	48.7	141	1	YPRO_ONEFU
7	149	48.1	2004	1	MOZ_HUMAN
8	148	47.7	487	1	EBN2_EBV
9	147.5	47.6	449	1	APG_BRANA
10	143	46.1	534	1	APG_ARATH
11	140.5	45.3	555	1	GPI_CHLRE
12	140	45.2	875	1	Y066_NPVOP
13	139.5	45.0	431	1	ACRO_RABIT
14	137.5	44.4	339	1	CSP_PLABE
15	137.5	44.4	347	1	CSP_PLABA
16	137	44.2	426	1	EXLP_TOBAC
17	136.5	44.0	565	1	MOTB_MOUSE
18	135.5	43.7	3110	1	HD_RAT
19	134.5	43.4	129	1	PARE_TRYBB
20	133.5	43.1	514	1	MEPD_MOUSE
21	133	42.9	440	1	G3PT_MOUSE
22	132	42.9	808	1	Y066_NPVAC
23	132.5	42.7	143	1	PAR1_TRYBB
24	132.5	42.7	145	1	PARC_TRYBB
25	132.5	42.7	979	1	REF1_HUMAN
26	132	42.6	282	1	GDA3_WHEAT
27	131	42.3	3144	1	HD_HUMAN
28	130.5	42.1	283	1	TONB_NEIGO
29	130.5	42.1	1206	1	FMI4_MOUSE
30	130.5	42.1	1468	1	FMN1_MOUSE
31	129.5	41.8	415	1	ACRO_PIG
32	129.5	41.8	722	1	Z219_HUMAN
33	128	41.3	224	1	Y091_NPVAC

34	127.5	41.1	791	1	Y046_UREPA
35	126.5	40.8	1139	1	KPCI_TIRE
36	126	40.6	1790	1	SEPA_EMENI
37	125.5	40.5	261	1	PRP2_MOUSE
38	125.5	40.5	443	1	HXA3_MOUSE
39	125	40.3	421	1	ACRO_HUMAN
40	125	40.3	3119	1	HD_MOUSE
41	124.5	40.2	293	1	HORI_HORVU
42	124.5	40.2	815	1	MK07_HUMAN
43	124.5	40.2	1362	1	BRD4_HUMAN
44	124	40.0	232	1	ACRL_HUMAN
45	123.5	39.8	296	1	PRP3_MOUSE

ALIGNMENTS

RESULT 1	PERT_BORPA	STANDARD;	PERT;	922 AA.
AC	P24328;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.70) (P.95).			
GN	PRN.			
OS	Bordetella parapertussis.			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=519;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN2591;			
RX	MEDLINE=91251771; PubMed=2041476;			
RA	Li L.J., Dougan G., Novotny P., Charles I.G.;			
RT	"P.70 pertactin, an outer-membrane protein from Bordetella			
RT	parapertussis: cloning, nucleotide sequence and surface expression in			
RT	Escherichia coli."			
RL	Mol. Microbiol. 5:409-417(1991).			
CC	-1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS			
CC	MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN			
CC	BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- SUBCELLULAR LOCATION: Outer membrane.			
CC	-1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)			
CC	CONCENTRATIONS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X54547; CAA38419.1; -			
DR	EMBL; A26124; CAA01786.1; -			
DR	EMBL; A19182; CAA01454.1; -			
DR	PIR; S15204; IS15204.			
DR	PIR; S14659; IS14659.			
DR	InterPro; IPF004899; Pertactin_sup.			
DR	InterPro; IPF003992; pertactin.			
DR	InterPro; IPF003991; pertactin_vir.			
DR	Pfam; PF03212; Pertactin; 1.			
DR	PRINTS; PR01482; PERTACTIN.			
DR	PRINTS; PR01484; PERTACTINFAMILY.			
KW	Outer membrane; Signal; Virulence; Repeat.			
FT	SIGNAL 1 34 POTENTIAL.			
FT	CHAIN 35 922 P.95.			
FT	CHAIN 35 647 PERTACTIN (P.70).			
FT	PROPEP 648 922 POTENTIAL.			
FT	SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN			

ID	YPRO_OWEFU	STANDARD;	PRT;	141 AA.
AC	P21260; P21261;			
DT	01-MAY-1991 {Rel. 18, Created}			
DT	01-MAY-1991 {Rel. 18, Last sequence update}			
DT	01-NOV-1995 {Rel. 32, Last annotation update}			
DE	Hypothetical proline-rich protein (Fragment).			
DE	Owenia fusiformis.			
OS	Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canaliculipata;			
OC				

CC Sabellida; Oweniidae; Owenia.
OX NCBI_TaxID=6347;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-90147742; PubMed=2105723;
RA Bakalara N., Collet J., Planells R., Thouveny Y., Fontes M.;
RT "Presence in invertebrate genomes of sequences characterized by the
RL repetition of the triplet CCGPine.";
RL Biochem. Biophys. Res. Commun. 166:66-73(1990).
DR PIR: A34043; A34043.
DR PIR: B34043; B34043.
KW Hypothetical protein; DNA-binding.
FT NON_TER 1
FT DOMAIN 9 58 POLY-PRO.
FT DOMAIN BIND 98 116 H-T-H MOTIF (POTENTIAL).
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15745 MW; B294E884D152BD89 CRC64;

Query Match 48.7%; Score 151; DB 1; Length 141;
Best Local Similarity 55.6%; Pred. No. 0.0011;
Matches 25; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 5 PAPKPAQPGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 49
DB 15 PPP 59

RESULT 7
MOZ_HUMAN STANDARD; PRT; 2004 AA.
ID Q92794;
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Monocytic leukemia zinc finger protein (zinc finger protein 220).
GN ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-96376968; PubMed=8782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horsman D., Mikelman F., Volinia S., Wamora A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -1- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U47742; AAC50662.1;
CC GenBank; HGNC:13013; ZNF220.
CC MIM; 601408;
CC InterPro; IPR001386; Histone_H1/H5.
CC InterPro; IPR002717; MOZ SAS.
CC InterPro; IPR001965; ZnF_PHD.
CC Pfam; PF00628; PHD; 2.
CC Pfam; PF01853; MOZ_SAS; 1.

DR SMART; SM00526; H15; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 MOZ-CBP.
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;

Query Match 48.1%; Score 149; DB 1; Length 2004;
Best Local Similarity 60.0%; Pred. No. 0.0094;
Matches 27; Conservative 1; Mismatches 13; Indels 4; Gaps 1;

OY 6 PAPKPAQPGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 46
DB 1651 PPP 1695

RESULT 8
EBN2_EBV STANDARD; PRT; 487 AA.
ID P12978;
AC P12978;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE EBNA-2 nuclear protein.
GN BYRF1.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuftnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SUBCELLULAR LOCATION, AND PHOSPHORYLATION.
RX MEDLINE-90266473; PubMed=2161150;
RA Petti L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
RN [3]
RP DOMAINS.
RX MEDLINE-91202599; PubMed=1850028;
RA Cohen J.I., Wang F., Kieff E.;
RT "Epstein-Barr virus nuclear protein 2 mutations define essential
RT domains for transactivation and transactivation.";
RL J. Virol. 65:2545-2554(1991).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. TRANSACTIVATES THE EXPRESSION
CC OF LMP-1.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC -1- PTM: PHOSPHORYLATED.

```

Query Match          47.6%  Score 147.5;  DB 1;  Length 449;
Best Local Similarity 44.4%  Pred. No. 0.0039;
Matches 28;  Conservative 4;  Mismatches 12;  Indels 19;  Gaps
QY 3 KVPKAPKAPGQPPQPPQ-----PPQPPQPPQPPQPPQ-----EAPAP 43
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 11 KPQKPPAPTPSCPQPPKPPKPPKPPAPTPSCPQPPKPPKPPAPGPPSP 70
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 44 QPP 46
      ||
Db 71 SPP 73

RESULT 10
APG-ARATH
ID      APG-ARATH      STANDARD;      PRT;      534 AA.
AC      P40602; Q93214; Q9LNT8;
DT      01-FEB-1995 (Rel. 31, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Anter-specific proline-rich protein APG precursor.
GN      APG OR APG240130 OR T20H2.9.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidaeaeurosid II; Brassicales; Brassicaceae; Arabidopsi.
OC      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94004980; PubMed=8401599;
RX      Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A., Draper J., Scott R.;
RT      "Gametophytic and sporophytic expression of an anther-specific Arabidopsis thaliana gene.";
RL      Plant J. 3:111-120(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia;
RX      MEDLINE=21016719; PubMed=11130712;
RX      Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White I.O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung J.M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koob H.L., Kremenetska I., Kurtz D.B., Kvan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A., Millischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Uutterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT      "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
RN      Nature 408:816-820(2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia;
RX      Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT      "RIKEN Arabidopsis full length cDNA clones (RAFTs) sequenced by the SFP consortium (Salk/Stanford/PGEc).";
RN      Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC      -1- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL TYPES IN THE ANTHR, ONLY IN MALE FERTILE PLANTS.
CC      -1- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN MATURATION.
CC      -1- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC      -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous

```

gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X60377; CAA42925.1; -
 CC EMBL; AC022472; AAF79900.1; ALT_SEQ.
 CC EMBL; AY058847; AAL24235.1; -
 CC PIR; S21961; S21961.
 CC InterPro; IPR001087; Lipase_GDSL.
 CC Pfam; PF00657; Lipase_GDSL; 1.
 CC PROSITE; PS01098; LIPASE_GDSL_SRR; 1.
 CC KW SIGNAL.
 CC FT CHAIN 1 35 POTENTIAL.
 CC FT ACT_SITE 36 534 ANTER-SPECIFIC PROLINE-RICH PROTEIN AFG.
 CC FT ACT_SITE 211 211 BY SIMILARITY.
 CC FT ACT_SITE 511 511 POTENTIAL.
 CC FT CONFLICT 77 77 S -> P (IN REF. 1).
 CC FT CONFLICT 141 141 E -> A (IN REF. 1).
 CC FT CONFLICT 325 325 E -> H (IN REF. 1).
 CC FT CONFLICT 325 325 E -> H (IN REF. 1).
 CC SQ SEQUENCE 534 AA; 58007 MW; BA851DC3CF7429DB CRC64;

Query Match 46.1%; Score 143; DB 1; Length 534;
 Best Local Similarity 52.9%; Pred. No. 0.0083;
 Matches 27; Conservative 3; Mismatches 13; Indels 8; Gaps 2;

OY 3 KVPAPKPAQPGQPQ-----PQPQPQPQPQPQPQPQPQPQPQPQPQPQP 46
 DB 98 KPPAPSP 147

RESULT 11
 GPI_CHLRE
 ID GPI_CHLRE STANDARD; PRT; 555 AA.
 AC Q9FPQ6; Q03927;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).
 GN GPI.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadales; Chlamydomonadales.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21159092; PubMed=11258910;
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J., Goodenough U.W.;
 RT "Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins.";
 RL Biochemistry 40:2978-2987(2001).
 RN [2]
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE-91017504; PubMed=1699225;
 RA Adair W.S., Apt K.E.;
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
 CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline) layer.
 CC -1- SUBUNIT: Associates with GP2 and GP3.
 CC -1- PTM: N-glycosylated and O-glycosylated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF309494; AAG45420.1; -
 CC EMBL; M58496; AAG69706.1; ALT_SEQ.
 CC GlycoSuiteDB; Q9FPQ6; -
 CC InterPro; IPR002965; P-rich_extensin.
 CC PRINTS; PRO1217; PRICHEXTENSIN.
 CC PRINTS; PRO1218; PSTEXTENSIN.
 CC KW Glycoprotein; Repeat; Signal.
 CC FT SIGNAL 1 29 POTENTIAL.
 CC FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
 CC FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSXP REPEATS.
 CC FT DOMAIN 259 279 POLY-PRO.
 CC FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 45.3%; Score 140.5; DB 1; Length 555;
 Best Local Similarity 54.3%; Pred. No. 0.012;
 Matches 25; Conservative 4; Mismatches 14; Indels 3; Gaps 2;

OY 4 VPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 47
 DB 245 VPPAP-PSAPSP 289

RESULT 12
 Y066_NPVOP
 ID Y066_NPVOP STANDARD; PRT; 875 AA.
 AC Q83949; Q65364; O10323;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Hypothetical 98.6 kDa protein (ORF71).
 OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OPMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=164623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97271300; PubMed=9126251;
 RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S., Rohrmann G.F.;
 RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.";
 RL Virology 229:381-399(1997).
 RN [2]
 RP SEQUENCE OF 1-807 FROM N.A.
 RX MEDLINE-96201426; PubMed=8609478;
 RA Ahrens C.H., Rohrmann G.F.;
 RT "The DNA polymerase and helicase genes of a baculovirus of Orgyia pseudotsugata.";
 RL J. Gen. Virol. 77:825-837(1996).
 RN [3]
 RP SEQUENCE OF 806-875 FROM N.A.
 RX MEDLINE-95343549; PubMed=7618274;
 RA Ahrens C.H., Carlson C., Rohrmann G.F.;
 RT "Identification, sequence, and transcriptional analysis of lef-3, a gene essential for Orgyia pseudotsugata baculovirus DNA replication.";
 RL Virology 210:372-382(1995).
 CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV AND LDMNPV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

FT DOMAIN 93 196 13 X 8 AA TANDEM REPEATS.


```
aps 2;
```


Db 1 GAKAPPAPKPAQPGQPPQP-----PQPQPEAPAPQPPAGRELSAA 42

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 28.0579 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754b-22
Perfect score: 310
Sequence: 1 GAKVPPAPKAPAPQPPQPPQ.....QPQPPAPAPQPPAGRELSAA 54

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	310	100.0	54	23	AAE16201 B. bronchiseptica
2	287	92.6	52	23	AAE16200 B. bronchiseptica
3	265	85.5	58	23	AAE16198 B. bronchiseptica
4	265	85.5	922	13	AAE25578 Bordetella parapertussis
5	265	85.5	922	23	AAE16185 Bordetella parapertussis
6	258	83.2	52	23	AAE16194 B. bronchiseptica
7	258	83.2	911	12	AAE14320 Pertactin antigen
8	258	83.2	911	13	AAE26503 prn proteins. Bor
9	258	83.2	911	23	AAE16183 Bordetella bronchi
10	256	82.6	56	23	AAE16197 B. bronchiseptica

11	255	82.3	922	12	AAE14321 Pertactin antigen
12	254	81.9	60	23	AAE16195 B. bronchiseptica
13	248	80.0	48	23	AAE16199 B. bronchiseptica
14	247.5	79.8	53	23	AAE16196 B. bronchiseptica
15	237.5	76.6	49	23	AAE16193 B. bronchiseptica
16	216	59.7	42	23	AAE16202 B. bronchiseptica
17	216	59.7	910	23	AAE16184 Bordetella pertussis
18	216	59.7	910	23	AAE17146 Bordetella pertussis
19	195.5	53.1	39	23	AAE16203 B. bronchiseptica
20	158	51.0	31	12	AAE11739 Peptide BB05 deriv
21	156	50.3	31	12	AAE11737 Peptide BB05 deriv
22	155.5	50.2	900	21	AAE4321 Human ORF2085
23	155.5	50.2	1217	22	ABG09876 Novel human diago
24	155.5	50.2	1239	22	ABG09877 Novel human diago
25	152	49.0	439	13	AAE28150 Sugar beet chitina
26	151	48.7	80	22	AAO04412 Human polypeptide
27	151	48.7	99	22	AAO02076 Human polypeptide
28	149	48.1	85	22	AAO04573 Human polypeptide
29	147.5	47.6	446	22	ABE70063 Drosophila melanog
30	144.5	46.6	88	22	AAO02124 Human polypeptide
31	143.5	46.3	85	22	AAO02036 Human polypeptide
32	143	46.1	350	23	ABE93838 Herbicidally activ
33	142	45.8	470	21	AAG49660 Arabidopsis thalia
34	141	45.5	406	22	ABG27250 Novel human diago
35	140.5	45.3	125	22	ABG17580 Novel human diago
36	139	44.8	358	21	AAG22949 Arabidopsis thalia
37	136.5	44.0	94	22	ABE67302 Drosophila melanog
38	136	43.9	76	23	ABE78537 Ser-Pro-Pro-Pro ge
39	136	43.9	261	23	ABE57327 Mouse ischaemic CO
40	135.5	43.7	272	10	AAE93560 Plasmodium berghei
41	135.5	43.7	329	23	AAE14305 Sugarcane proline
42	135	43.5	526	21	AAE12717 Streptococcus pneu
43	135	43.5	1749	22	AAE72314 Glutamic acid rich
44	134.5	43.4	1222	21	AAG36453 Arabidopsis thalia
45	134.5	43.4	1257	21	AAG36452 Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAE16201
ID AAE16201 standard; peptide; 54 AA.
XX AAE16201;
AC AAE16201;
XX 26-MAR-2002 (first entry)
DT B. bronchiseptica strain II-9 pertactin outer membrane protein region II.
XX Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX therapy; antibiotic; antibacterial; region II.
OS Bordetella bronchiseptica.
XX WO200190143-A2.
PN 29-NOV-2001.
PD 23-MAY-2001; 2001WO-EP06457.
PF 25-MAY-2000; 2000US-206969P.
PR (INSP) INST PASTEUR.
XX Guiso-maclouf N, Boursaux-eude C;
XX WPI; 2002-097639/13.
XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

[illegible]


```
FT Region 570..589
FT /label= Repeat_region
FT Peptide 260..262
FT /label= RGD_tripeptide
FT Peptide 701..703
FT /label= RGD_tripeptide
XX
XX WO9217587-A.
XX
XX PD 15-OCT-1992.
XX
XX PF 27-MAR-1992; 92WO-GB00561.
XX
XX PR 27-MAR-1991; 91GB-0006568.
XX
XX PA (WELL ) WELLCOME FOUND LTD.
XX
XX PI Charles IG;
XX
XX DR WPI; 1992-366258/44.
XX N-PSDB; AAQ34566.
XX
XX DNA encoding a Bordetella bronchiseptica protein - used for
XX obtaining vaccines for preventing respiratory diseases, partic.
XX atrophic rhinitis in pigs
XX
XX PS Claim 1; Fig 1; 28pp; English.
XX
XX The sequence given is the P.94 antigen from B. bronchiseptica. The
XX P.68 antigen is formed by alternative cleavage of this protein.
XX P.68 is an outer membrane protein with a molecular weight of 68 kD
XX which is associated with protection of piglets against atrophic
XX rhinitis. The DNA sequence encoding these proteins was derived by
XX standard recombinant DNA techniques using P.68 probes to isolate the
XX entire P.94 sequence.
XX
XX SQ Sequence 911 AA;
XX
XX Query Match 83.28; Score 258; DB 13; Length 911;
XX Best Local Similarity 89.18; Pred. No. 4.8e-12;
XX Matches 49; Conservative 0; Mismatches 2; Indels 4; Gaps 2;
XX
XX QY 1 GAKVPPAPKPAQPGQP-PQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 54
XX ||| ||||||||||| ||||||||||| || |||||||||||||||
XX Db 559 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 610
XX
XX RESULT 9
XX AAE16183
XX ID AAE16183 standard; Protein: 911 AA.
XX
XX AC AAE16183;
XX
XX XX 26-MAR-2002 (first entry)
XX
XX Bordetella bronchiseptica pertactin outer membrane protein, p.68.
XX
XX DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX KW therapy; antibiotic; antibacterial; p.68.
XX
XX OS Bordetella bronchiseptica.
XX
XX FH Key Location/Qualifiers
XX FT 254..299
XX FT /note= "pertactin region I"
XX FT Region 559..610
XX FT /note= "pertactin region II"
XX
XX PN WO200190143-A2.
XX
XX PD 29-NOV-2001.
XX
XX XX 23-MAY-2001; 2001WO-EP06457.
XX
XX PF Polypeptides containing polymorphisms of the repeated regions of
XX pertactin in Bordetella species, useful in immunogenic compositions for
XX treating infections caused by Bordetella and in diagnostic methods -
```

```
XX
XX 25-MAY-2000; 2000US-206969P.
XX (INSP ) INST PASTEUR.
XX
XX Guiso-maclouf N, Boursaux-eude C;
XX
XX WPI; 2002-097639/13.
XX N-PSDB; AAD26440.
XX
XX Polypeptides containing polymorphisms of the repeated regions of
XX pertactin in Bordetella species, useful in immunogenic compositions for
XX treating infections caused by Bordetella and in diagnostic methods -
XX
XX Disclosure; Page 28; 47pp; English.
XX
XX The present invention relates to Bordetella bronchiseptica pertactin
XX (outer membrane protein) or their fragments. Pertactin (PRN) is used as
XX vaccine. Pertactin antibody is useful for treating Bordetella infections
XX and used to detect Bordetella antigens in biological preparations or in
XX purifying corresponding proteins, glycoproteins or their mixtures when
XX used in affinity chromatographic columns. Pertactin is useful as antigens
XX to identify antibodies to Bordetella in materials such as human or other
XX animal tissue and human or other animal cells, as well as biological
XX fluids, such as human or other animal body fluids, including human sera,
XX and to determine the concentration of Ab in those materials. Thus the
XX antigens can be used for qualitative or quantitative determination of
XX Bordetella in a material. The present sequence is B. bronchiseptica
XX pertactin outer membrane protein, p.68.
XX
XX SQ Sequence 911 AA;
XX
XX Query Match 83.28; Score 258; DB 23; Length 911;
XX Best Local Similarity 89.18; Pred. No. 4.8e-12;
XX Matches 49; Conservative 0; Mismatches 2; Indels 4; Gaps 2;
XX
XX QY 1 GAKVPPAPKPAQPGQP-PQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 54
XX ||| ||||||||||| ||||||||||| || |||||||||||||||
XX Db 559 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 610
XX
XX RESULT 10
XX AAE16197
XX ID AAE16197 standard; peptide: 56 AA.
XX
XX AC AAE16197;
XX
XX XX 26-MAR-2002 (first entry)
XX
XX B. bronchiseptica strain II-5 pertactin outer membrane protein region II.
XX
XX DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
XX KW therapy; antibiotic; antibacterial; region II.
XX
XX OS Bordetella bronchiseptica.
XX
XX PN WO200190143-A2.
XX
XX PD 29-NOV-2001.
XX
XX XX 23-MAY-2001; 2001WO-EP06457.
XX
XX PF Polypeptides containing polymorphisms of the repeated regions of
XX pertactin in Bordetella species, useful in immunogenic compositions for
XX treating infections caused by Bordetella and in diagnostic methods -
```

PD 17-OCT-1991.

The present invention relates to Bordetella bronchiseptica pertactin (outer membrane protein) or their fragments. Pertactin (PRN) is used as

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.0494 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-21

Perfect score: 298

Sequence: 1 GAKVPPAPKAPQPGQPQ.....QPQEPAPQPPAGRELSAA 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	259	86.9	911	1	PERT_BORBR
2	256	85.9	922	1	PERT_BORPA
3	217	72.8	910	1	PERT_BORPE
4	147	49.3	485	1	SSGP_VOLCA
5	145.5	48.8	3164	1	TEGU_HSV11
6	145	48.7	141	1	YPRO_OWEFU
7	144.5	48.5	449	1	APC_BRANA
8	142	47.7	2004	1	MOZ_HUMAN
9	141	47.3	487	1	EBN2_EBV
10	133.5	44.8	426	1	EXLP_TOBAC
11	132.5	44.5	431	1	ACRO_RABIT
12	132.5	44.5	534	1	APG_ARATH
13	131.5	44.1	3110	1	HD_RAT
14	131	44.0	440	1	G3PT_MOUSE
15	131	44.0	875	1	Y066_NPVAC
16	129.5	43.5	339	1	CSP_PLABE
17	129.5	43.5	347	1	CSP_PLABA
18	129.5	43.5	555	1	GPI_CHLRE
19	128	43.0	3144	1	HD_HUMAN
20	127.5	42.8	565	1	MOT8_MOUSE
21	127	42.6	282	1	GDA3_WHEAT
22	127	42.6	808	1	Y066_NPVAC
23	126	42.3	421	1	ACRO_HUMAN
24	126	42.3	3119	1	HD_MOUSE
25	125.5	42.1	979	1	RF11_HUMAN
26	125.5	42.1	1790	1	SEPA_EMENI
27	125	41.9	224	1	Y091_NPVAC
28	125	41.9	232	1	ACRL_HUMAN
29	125	41.9	415	1	ACRO_PIG
30	125	41.9	1362	1	BRD4_HUMAN
31	124.5	41.8	129	1	PARRB_TRYBB
32	124.5	41.8	722	1	2219_HUMAN
33	124.5	41.8	1206	1	FM14_MOUSE

34	124.5	41.8	1468	1	FMN1_MOUSE
35	124	41.6	296	1	GDA6_WHEAT
36	123.5	41.4	143	1	PARL_TRYBB
37	123.5	41.4	145	1	PARC_TRYBB
38	123.5	41.4	267	1	EXTN_MAIZE
39	122.5	41.1	1139	1	KPC1_TIRE
40	121.5	40.8	443	1	HXA2_MOUSE
41	121	40.6	2715	1	TRX2_HUMAN
42	120.5	40.4	261	1	PRP2_MOUSE
43	120.5	40.4	283	1	TONB_NEIGO
44	120.5	40.4	296	1	PRP3_MOUSE
45	120.5	40.4	514	1	MEFD_MOUSE

ALIGNMENTS

RESULT 1

PERT_BORBR
ID PERT_BORBR STANDARD; PRT; 911 AA.
AC Q03035;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.68) (P.94).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.
RC STRAIN=CN7531;
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica.";
RL J. Gen. Microbiol. 138:1697-1705(1992).
CC -!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATIONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54815; CAA38584.1; -
CC EMBL; A19180; CAA01453.1; -
CC PIR; A47675; A47675.
CC InterPro: IPR004899; Pertactin sup.
CC InterPro: IPR003992; pertactin.
CC InterPro: IPR003991; pertactin_vir.
CC Pfam: PF03212; Pertactin; 1.
CC PRINTS; PR01482; PERTACTIN.
CC PRINTS; PR01484; PERTACTINFAMLY.
CC Outer membrane; Signal; Virulence; Repeat.
CC SIGNAL 1 34
CC CHAIN 35 911 P.94
CC CHAIN 35 712 PERTACTIN (P.68).
CC PROPEP 713 911 POTENTIAL.
CC SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
CC ADHESION TO VARIOUS EUKARYOTIC CELL
CC LINES).

[illegible]

Query Match 48.5%; Score 144.5; DB 1; Length 449;
Best Local Similarity 48.3%; Pred. No. 0.0049;
Matches 28; Conservative 3; Mismatches 12; Indels 15; Gaps 2

QY 3 KVPKAPKAPGPPQPQQP-----PPQPPQPPQPPAPAPAPQPPAG 46
| | | | | | | | | | | | | | | | | | | | | |
DB 11 KEQAPPPAPTSPCPPPKPKPPKPAPPAPTSPCPQPPKPPQPKP-PPAPGPSRPG 67
| | | | | | | | | | | | | | | | | | | | | |

RESULT 8

MOZ_HUMAN STANDARD; PRT; 2004 AA.

ID AC Q92794;
AD 15-JUL-1998 (Rel. 36, Created)
DT DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE Monocytic leukemia zinc finger protein (zinc finger protein 220).
DE ZNF220 OR MOZ.
GN GN Homo sapiens (Human).
OS OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Civin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horstman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11:P13) CHROMOSOMAL
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a colla-
between the Swiss Institute of Bioinformatics and the EMBL out-
the European Bioinformatics Institute. There are no restrictions
use by non-profit institutions as long as its content is
modified and this statement is not removed. Usage by and for com-
entities requires a license agreement (See http://www.isb-sib.ch/a/
or send an email to license@isb-sib.ch).

EMBL; U47742; AAC50662.1; --
Genew; HGNC:13013; ZNF220.
MIM; 601408; --
DR DR InterPro; IPR001386; Histone_H1/H5.
DR DR InterPro; IPR002717; MOZ_SAS.
DR DR InterPro; IPR001965; ZnF_PHD.
DR DR Pfam; PF00628; PHD; 2.
DR DR Pfam; PF01853; MOZ_SAS; 1.
DR DR SMART; SM00526; H15; 1.
DR DR SMART; SM00249; PHD; 2.
DR DR PROSITE; PS01359; ZF_PHD_1; 1.
DR DR PROSITE; PS00016; ZF_PHD_2; 2.
KW KW Proto-oncogene. Chromosomal translocation; zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.

DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
GN ACR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Testis;
RX MEDLINE=9436861; PubMed=8086468;
RA Richardson R.T., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
RT preproacrosin-related cDNA."
RL Biochim. Biophys. Acta 1219:215-218(1994).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOEA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
CC ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U05204; AAA61630.1; --
DR HSP: P00763; LDPO.
DR MEROPS: S01.223;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS02440; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_SER; 1.
DR PROSITE: PS00135; TRYPSIN_HIS; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16
FT CHAIN 17 431
FT CHAIN 17 39 ACROSIN
FT CHAIN 40 431 ACROSIN LIGHT CHAIN (BY SIMILARITY).
FT PROPEP 7 431 PRO-RICH.
FT DOMAIN 40 288 SERINE PROTEASE.
FT DISULFID 22 152 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 160 INTERCHAIN (BY SIMILARITY).
FT DISULFID 71 87 BY SIMILARITY.
FT DISULFID 175 244 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 234 264 BY SIMILARITY.
FT CARBOHYD 19 19 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC...) (POTENTIAL).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 431 AA; 46422 MW; 1C015A4E0BC0C668 CRC64;

Query Match 44.5%; Score 132.5; DB 1; Length 431;
Best Local Similarity 45.5%; Pred. No. 0.025;
Matches 25; Conservative 0; Mismatches 19; Indels 11; Gaps 1;

QY 2 AKVPPAPKAPQPGFQ-----PPQPPFPQPPQPPQPPQPPQPPA 45
DB 325 ASGPPPHPHPHPHPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 379

RESULT 12
APG_ARATH STANDARD; PRT; 534 AA.
ID AC P40602; Q93214; Q9LNT8;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Anter-specific proline-rich protein APG precursor.
GN APG OR A1G20130 OR T20H2.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94004980; PubMed=8401599;
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA Draper J., Scott R.;
RT "Gametophytic and sporophytic expression of an anther-specific
RT Arabidopsis thaliana gene."
RL Plant J. 3:111-120(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buchler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kvan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGSC)."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC TYPES IN THE ANTER, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X60377; CAA42925.1; --


```
CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT DAY 20 IN POST-MEIOTIC
CC GERM CELLS. LEVELS INCREASE UNTIL DAY 24 AND THEN REMAIN CONSTANT
CC DURING MATURITY.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60978; AAA53033.1; -
DR EMBL; U09564; AAA80276.1; -
DR HSSP; P56649; LDSS.
DR MGD; MGI:95653; Gapds.
DR InterPro; IPR000173; GAP_dhhydrogenase.
DR Pfam; PF00044; gpdh; 1.
DR Pfam; PF02800; gpdh.C; 1.
DR PRINTS; PR00078; G3PDHGRGNASE.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT DOMAIN 21 41 CYS/PRO-RICH.
FT DOMAIN 54 73 POLY-PRO.
FT DOMAIN 84 100 POLY-PRO.
FT BINDING 256 256 GLYCERALDEHYDE 3-PHOSPHATE (BY
FT SIMILARITY).
FT ACT_SITE 283 283 ACTIVATES THIOL GROUP DURING CATALYSIS
FT CONFLICT 33 34 (BY SIMILARITY).
FT CONFLICT 43 43 MISSING (IN REF. 2).
FT CONFLICT 43 43 L -> V (IN REF. 2).
SQ SEQUENCE 440 AA; 47657 MW; 05FF0A093D1ABD9C CRC64;

Query Match 44.0%; Score 131; DB 1; Length 440;
Best Local Similarity 49.1%; Pred. No. 0.031;
Matches 27; Conservative 1; Mismatches 17; Indels 10; Gaps 2;

Qy 5 PPAPKAPQGPQPQPQPQ-----PPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 50
Db 56 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 109

RESULT 15
Y066_NPVOP STANDARD; PRT; 875 AA.
AC Q83949; O65364; O10323.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 98.6 kDa protein (ORF71).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97271300; PubMed-9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
RN [2]
RP SEQUENCE OF 1-807 FROM N.A.
RX MEDLINE-96201426; PubMed-8609478;
RA Ahrens C.H., Rohrmann G.F.;
RT "The DNA polymerase and helicase genes of a baculovirus of Orgyia
RT pseudotsugata."
RL J. Gen. Virol. 77:825-837(1996).
RN [3]
```

```
RP SEQUENCE OF 806-875 FROM N.A.
RX MEDLINE-95343549; PubMed-7618274;
RA Ahrens C.H., Carlson C., Rohrmann G.F.;
RT "Identification, sequence, and transcriptional analysis of lef-3, a
RT gene essential for Orgyia pseudotsugata baculovirus DNA
RT replication."
RL Virology 210:372-382(1995).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACNMPV AND LDMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59070.1; -
DR EMBL; U39145; AAB04047.1; -
DR EMBL; D45397; BAA08236.1; -
DR HSSP; P01100; LFOS.
KW Hypothetical protein.
FT DOMAIN 86 91 POLY-PRO.
SQ SEQUENCE 875 AA; 98603 MW; F03DB1B430D33C42 CRC64;

Query Match 44.0%; Score 131; DB 1; Length 875;
Best Local Similarity 51.9%; Pred. No. 0.052;
Matches 28; Conservative 1; Mismatches 11; Indels 14; Gaps 3;

Qy 5 PPAPKP-----APQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 44
Db 86 PPPPQPFYGYWFPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 139

Search completed: May 7, 2003, 16:48:14
Job time : 7.0494 secs
```



```

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin.
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B14;
RA King A., Berbers G., Hoogerhout P., Oirschot van H.F., Knipping K.,
RA Mooi F.R.;
RT "Role of the polymorphic region of the B.pertussis protein pertactin
RT in immunity: implications for the design of pertussis vaccines.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ245927; CAB82515.1; -.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PRACINFAMILY.
DR SEQUENCE 911 AA; 94093 MW; DF531A9EB4383A32 CRC64;
SQ
Query Match 86.9%; Score 259; DB 2; Length 911;
Best Local Similarity 92.5%; Pred. No. 2.3e-14;
Matches 49; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

OY 1 GAKVPPAPKAPQGPQP-PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 52
DB 559 GAKAPPAPKAPQGPQGPQGPQPQPQPQP-PQRPQPEAPAPQPPAGRELSAA 610

RESULT 10
ID Q9K5G9 PRELIMINARY; PRT; 111 AA.
AC Q9K5G9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin (P.68) (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BBCH;
RX MEDLINE-20359389; PubMed-10899896;
RA Boursaux-Eude C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella
RT pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250087; CAB76441.1; -.
DR InterPro; IPR004899; Pertact_sup.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR NON_TER 1
FT NON_TER 111
FT SEQUENCE 111 AA; 11416 MW; 5140659692808F8E CRC64;
SQ
Query Match 86.2%; Score 257; DB 2; Length 111;
Best Local Similarity 86.0%; Pred. No. 5.8e-15;
Matches 49; Conservative 0; Mismatches 2; Indels 6; Gaps 2;

OY 1 GAKVPPAPKAPQGPQP-PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 52
DB 21 GAKAPPAPKAPQGPQGPQGPQPQPQPQP-PQRPQPEAPAPQPPAGRELSAA 76

us-09-855-754b-21.rspt
```

```

RESULT 11
OY Q9KJX9 PRELIMINARY; PRT; 122 AA.
AC Q9KJX9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin (Fragment).
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-16039;
RA Keil D.J., Fenwick B.;
RT "Variability in the pertactin genes of Bordetella bronchiseptica
RT isolates includes regions coding for the GXXP and PGP families of
RT amino acid repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156773; AAF82397.1; -.
DR InterPro; IPR004899; Pertact_sup.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR NON_TER 1
FT NON_TER 122
FT SEQUENCE 122 AA; 12395 MW; 8ED00966A40FF994 CRC64;
SQ
Query Match 86.2%; Score 257; DB 2; Length 122;
Best Local Similarity 86.0%; Pred. No. 6.3e-15;
Matches 49; Conservative 0; Mismatches 2; Indels 6; Gaps 2;

OY 1 GAKVPPAPKAPQGPQP-PQPQPQPQPQPQPQPQPQPQPQPAPQPPAGRELSAA 52
DB 42 GAKAPPAPKAPQGPQGPQGPQPQPQPQP-PQRPQPEAPAPQPPAGRELSAA 97

RESULT 12
OY Q9ALQ2 PRELIMINARY; PRT; 252 AA.
AC Q9ALQ2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AVCC19395;
RX MEDLINE-21117018; PubMed-11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella
RT bronchiseptica Pertactin.";
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL; AY007265; AAG38441.1; -.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01217; PRICHEXTENS.
DR NON_TER 1
FT NON_TER 252
FT SEQUENCE 252 AA; 26107 MW; 368C142508D77057 CRC64;
SQ
Query Match 86.2%; Score 257; DB 2; Length 252;
Best Local Similarity 86.0%; Pred. No. 1.1e-14;
Matches 49; Conservative 0; Mismatches 2; Indels 6; Gaps 2;
```

FT	NON_TER	115	115
----	---------	-----	-----

Job time : 27.5753 secs


```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-886-29

```

[illegible]

```

RESULT 8
US-09-041-886-30
; Sequence 30, Application US/09041886
; Patent NO. 6235872
; GENERAL INFORMATION:
; APPLICANT: Brødesen, Dale E.
; APPLICANT: Rabizadeh, Shartoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370-La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122

```

ZIF: 52122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/041,886
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 2626
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 30:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 552 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-041-886-30

Query Match 43.0%; Score 128; DB 4; Length 552;
Best Local Similarity 48.3%; Pred. NO. 0.00022;
Matches 28; Conservative 1; Mismatches 15; Indels

```

QY      6 PAPKAPAGPQPQQPPQPQQPPQPQQPEAPAPQPPAG-----RELSA 51
| | | | | | | | | | | | | | | | : | | | |
DB    41 PPPPPPPPPQLGQPQQPPQAQLLPQFQPPP--ppppppPGPAVAEPLHRPKELSA 96
| | | | | | | | | | | | | | | | : | | | |

RESULT 9
US-09-041-886-31
; Sequence 31, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Prospoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886

```

```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
UNUS-09-041-886-31

```

Query Match 43.0%; Score 128; DB 4; Length 589;
Best Local Similarity 48.3%; Pred. No. 0.00023;
Matches 28: Conservative 1; Mismatches 15; Indels 14; Gaps 2;

[illegible]

RESULT 10
US-08-246-982A-6
; Sequence 6, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: Macdonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

```

1  ZIP: 20003
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: Patent In Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/453,265
16
17 FILING DATE: 30-MAY-1995
18
19 CLASSIFICATION: 514
20
21 ATTORNEY/AGENT INFORMATION:
22
23 NAME: Ludwig, Steven R.
24
25 REGISTRATION NUMBER: 36,203
26
27 REFERENCE/DOCKET NUMBER: 0609.3880003
28
29 TELECOMMUNICATION INFORMATION:
30
31 TELEPHONE: (202) 371-2600
32
33 TELEFAX: (202) 371-2540
34
35 INFORMATION FOR SEQ ID NO: 6:
36
37 SEQUENCE CHARACTERISTICS:
38
39 LENGTH: 3144 amino acids
40
41 TYPE: amino acid
42
43 TOPOLOGY: linear
44

```

```

RESULT 13
US-08-556-419-21
; Sequence [21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: ROSS, Christopher.
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Shapp, Alan

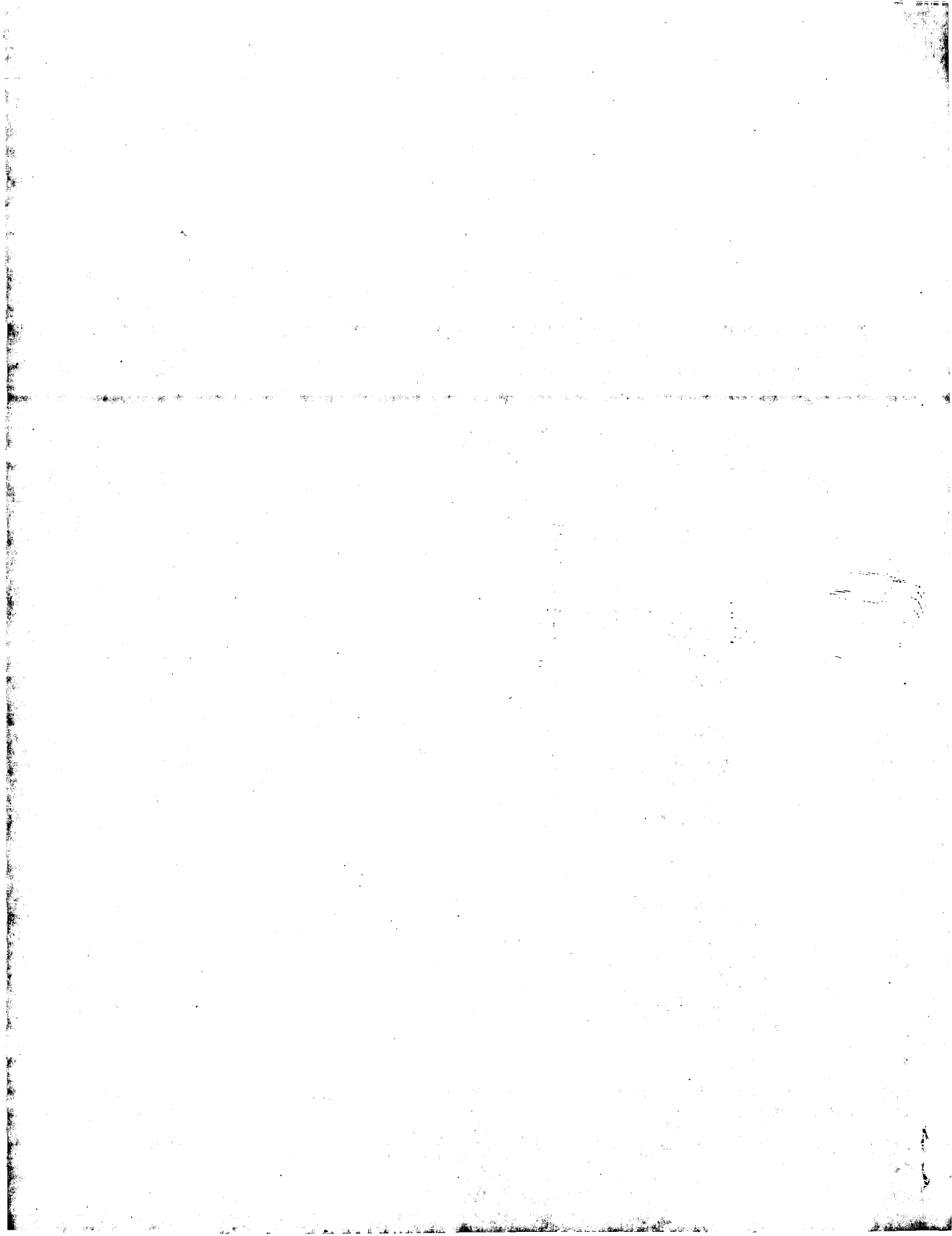
```


Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	298	100.0	52	9	US-09-855-754-21	Sequence 21, Appl
2	287	96.3	54	9	US-09-855-754-22	Sequence 22, Appl
3	259	86.9	52	9	US-09-855-754-15	Sequence 15, Appl
4	259	86.9	911	9	US-09-855-754-4	Sequence 4, Appl
5	257	86.2	56	9	US-09-855-754-18	Sequence 18, Appl
6	256	85.9	58	9	US-09-855-754-19	Sequence 19, Appl
7	256	85.9	922	9	US-09-855-754-6	Sequence 6, Appl
8	249	83.6	48	9	US-09-855-754-20	Sequence 20, Appl
9	242.5	81.4	59	9	US-09-855-754-16	Sequence 16, Appl
10	238.5	80.0	49	9	US-09-855-754-14	Sequence 14, Appl
11	236	79.2	52	9	US-09-855-754-17	Sequence 17, Appl
12	217	72.8	42	9	US-09-855-754-23	Sequence 23, Appl
13	217	72.8	910	9	US-09-855-754-5	Sequence 5, Appl
14	196.5	65.9	39	9	US-09-855-754-24	Sequence 24, Appl
15	140	47.0	827	9	US-10-171-384-3	Sequence 3, Appl
16	129.5	43.5	641	10	US-09-755-272-160	Sequence 160, Appl
17	128	43.0	1543	10	US-09-904-987-7	Sequence 7, Appl
18	127.5	41.8	1134	9	US-10-081-873-50	Sequence 50, Appl
19	125	42.9	68	9	US-10-077-584-9	Sequence 9, Appl

GENERAL INFORMATION:
; APPLICANT: The Brigham & Women's Hospital, Inc.
; APPLICANT: Beier, David
; APPLICANT: Herron, Bruce
; APPLICANT: Rao, Cherie
; TITLE OF INVENTION: p53 Binding Protein-Related Protein in Cardiomyopathy
; FILE REFERENCE: 81994/275368
; CURRENT APPLICATION NUMBER: US/10/171,384
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-384-3

Query Match 47.0%; Score 140; DB 9; Length 827;
Best Local Similarity 62.5%; Pred. No. 0.0082;
Matches 25; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 5 PPAPKPAQPGPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 44
|| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 PPKLQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 441

Search completed: May 7, 2003, 17:30:54
Job time : 15.5911 secs



Result No.	Score	Query Match	Length	ID	Description
1	259	86.9	911	2	A17675
2	256	85.9	922	2	S15204
3	217	72.8	910	2	A32560
4	162.5	54.5	430	2	JC2301
5	149	50.0	1494	2	T14355
6	148	49.7	209	2	A48232
7	148	49.7	599	2	T10798
8	147	49.3	485	2	A33647
9	145.5	48.8	3164	1	WBEH6
10	145	48.7	141	2	A34043
11	144.5	48.5	449	2	S16748
12	144.5	48.5	1952	2	T48814
13	143	48.0	383	2	AE2295
14	141	47.3	487	2	S12442
15	140	47.0	544	2	T17547
16	137.5	46.1	534	2	S21961
17	135.5	45.5	196	2	B48232
18	135	45.3	439	2	S19339
19	134.5	45.1	165	2	T24470
20	134.5	45.1	165	2	T07907
21	133.5	44.8	426	2	J01696
22	133.5	44.8	1474	2	T18281
23	133	44.6	320	2	E23364
24	132.5	44.5	431	2	S47538
25	132.5	44.5	464	2	S22697
26	132.5	44.5	1137	2	A86335
27	131	44.0	440	2	T10340
28	131	44.0	875	2	B06381
29	130	43.6	494	2	T96534
30	130	43.6	494	2	T96534
31	130	43.6	494	2	T96534
32	130	43.6	494	2	T96534
33	130	43.6	494	2	T96534
34	130	43.6	494	2	T96534
35	130	43.6	494	2	T96534
36	130	43.6	494	2	T96534
37	130	43.6	494	2	T96534
38	130	43.6	494	2	T96534
39	130	43.6	494	2	T96534
40	130	43.6	494	2	T96534
41	130	43.6	494	2	T96534
42	130	43.6	494	2	T96534
43	130	43.6	494	2	T96534
44	130	43.6	494	2	T96534
45	130	43.6	494	2	T96534
46	130	43.6	494	2	T96534
47	130	43.6	494	2	T96534
48	130	43.6	494	2	T96534
49	130	43.6	494	2	T96534
50	130	43.6	494	2	T96534
51	130	43.6	494	2	T96534
52	130	43.6	494	2	T96534
53	130	43.6	494	2	T96534
54	130	43.6	494	2	T96534
55	130	43.6	494	2	T96534
56	130	43.6	494	2	T96534
57	130	43.6	494	2	T96534
58	130	43.6	494	2	T96534
59	130	43.6	494	2	T96534
60	130	43.6	494	2	T96534
61	130	43.6	494	2	T96534
62	130	43.6	494	2	T96534
63	130	43.6	494	2	T96534
64	130	43.6	494	2	T96534
65	130	43.6	494	2	T96534
66	130	43.6	494	2	T96534
67	130	43.6	494	2	T96534
68	130	43.6	494	2	T96534
69	130	43.6	494	2	T96534
70	130	43.6	494	2	T96534
71	130	43.6	494	2	T96534
72	130	43.6	494	2	T96534
73	130	43.6	494	2	T96534
74	130	43.6	494	2	T96534
75	130	43.6	494	2	T96534
76	130	43.6	494	2	T

R; Bakalara, N.; Collet, J.; Planells, R.; Thouveny, Y.; Fontes, M.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 27.0187 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-21

Perfect score: 298

Sequence: 1 GAKVPPAPKPPQPPQ.....QPQEPAPQPPAGRELSAA 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	298	100.0	52	23	AAE16200
2	287	96.3	54	23	AAE16201
3	259	86.9	52	23	AAE16194
4	259	86.9	911	12	AAE14320
5	259	86.9	911	13	AAE26503
6	259	86.9	911	23	AAE16183
7	257	86.2	56	23	AAE16197
8	256	85.9	58	23	AAE16198
9	256	85.9	922	13	AAE25578
10	256	85.9	922	23	AAE16185

11	255	85.6	60	23	AAE16195
12	249	83.6	48	23	AAE16199
13	248.5	83.4	53	23	AAE16196
14	246	82.6	922	12	AAE14321
15	238.5	80.0	49	23	AAE16193
16	217	72.8	42	23	AAE16202
17	217	72.8	910	23	AAE16184
18	217	72.8	910	23	AAE17146
19	196.5	65.9	39	23	AAE16203
20	159	53.4	31	12	AAE11739
21	157	52.7	31	12	AAE11737
22	148.5	49.8	900	21	AAE42321
23	148.5	49.8	1217	22	ABG09876
24	148.5	49.8	1239	22	ABG09877
25	145	48.7	80	22	AAO04412
26	145	48.7	99	22	ABO02076
27	144.5	48.5	446	22	ABO70063
28	143	48.0	85	22	AAO04573
29	141	47.3	88	22	AAO02124
30	137.5	46.1	85	22	AAO02036
31	136.5	45.8	470	21	AAE49660
32	136	45.6	350	23	ABO93838
33	136	45.6	406	22	ABG27250
34	135	45.3	439	13	AAE28150
35	134.5	45.1	358	21	AAE22949
36	133.5	44.8	125	22	ABG17580
37	131.5	44.1	94	22	ABG67302
38	131	44.0	44	22	AAO04929
39	130.5	43.8	325	22	ABG21919
40	129.5	43.5	233	18	AAE14572
41	129.5	43.5	272	10	AAE93560
42	129.5	43.5	641	19	AAE61217
43	129.5	43.5	641	23	AAE54636
44	129.5	43.5	745	21	AAE81652
45	128.5	43.1	76	23	ABE78537

ALIGNMENTS

RESULT 1

AAE16200

ID AAE16200 standard; peptide; 52 AA.

XX

AC AAE16200;

XX

DT 26-MAR-2002 (first entry)

XX

XX B. bronchiseptica strain II-8 pertactin outer membrane protein region II.

DE Pertactin; PFN; outer membrane protein; vaccine; Bordetella infection;

XX therapy; antibiotic; antibacterial; region II.

KW Bordetella bronchiseptica.

XX WO200190143-A2.

OS 29-NOV-2001.

XX 23-MAY-2001; 2001WO-EP06457.

XX 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of

PT pertactin in Bordetella species, useful in immunogenic compositions for

PT treating infections caused by Bordetella and in diagnostic methods

XX

```

PS Claim 26; Fig 1c; 47pp; English.
CC The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region II.
XX
SQ Sequence 52 AA;

Query Match 100.0%; Score 298; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKVPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52
DB 1 GAKVPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52

RESULT 2
AAE16201
ID AAE16201 standard; peptide; 54 AA.
AC AAE16201;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-9 pertactin outer membrane protein region II.
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969p.
XX
PA (INSP ) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969p.
XX
PA (INSP ) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
PS Claim 26; Fig 1c; 47pp; English.
XX
XX The present invention relates to Bordetella bronchiseptica pertactin
XX (outer membrane protein) or their fragments. Pertactin (PRN) is used as
XX vaccine. Pertactin antibody is useful for treating Bordetella infections
XX and used to detect Bordetella antigens in biological preparations or in
XX purifying corresponding proteins, glycoproteins or their mixtures when
XX used in affinity chromatographic columns. Pertactin is useful as antigens
XX to identify antibodies to Bordetella in materials such as human or other
XX animal tissue and human or other animal cells, as well as biological
XX fluids, such as human or other animal body fluids, including human sera,
XX and to determine the concentration of Ab in those materials. Thus the
XX antigens can be used for qualitative or quantitative determination of
XX Bordetella in a material. The present sequence is B. bronchiseptica
XX pertactin outer membrane protein region II.
XX
SQ Sequence 52 AA;

Query Match 100.0%; Score 298; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKVPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52
DB 1 GAKVPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52

RESULT 3
AAE16194
ID AAE16194 standard; peptide; 52 AA.
AC AAE16194;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE B. bronchiseptica strain II-2 pertactin outer membrane protein region II.
KW Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;
KW therapy; antibiotic; antibacterial; region II.
XX
OS Bordetella bronchiseptica.
XX
PN WO200190143-A2.
XX
PD 29-NOV-2001.
XX
PF 23-MAY-2001; 2001WO-EP06457.
XX
PR 25-MAY-2000; 2000US-206969p.
XX
PA (INSP ) INST PASTEUR.
XX
PI Guiso-maclouf N, Boursaux-eude C;
XX
DR WPI; 2002-097639/13.
XX
PT Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods
XX
PS Claim 26; Fig 1c; 47pp; English.
XX
XX The present invention relates to Bordetella bronchiseptica pertactin
XX (outer membrane protein) or their fragments. Pertactin (PRN) is used as
XX vaccine. Pertactin antibody is useful for treating Bordetella infections
XX and used to detect Bordetella antigens in biological preparations or in
XX purifying corresponding proteins, glycoproteins or their mixtures when
XX used in affinity chromatographic columns. Pertactin is useful as antigens
XX to identify antibodies to Bordetella in materials such as human or other
XX animal tissue and human or other animal cells, as well as biological
XX fluids, such as human or other animal body fluids, including human sera,
XX and to determine the concentration of Ab in those materials. Thus the
XX antigens can be used for qualitative or quantitative determination of
XX Bordetella in a material. The present sequence is B. bronchiseptica
XX pertactin outer membrane protein region II.
XX
SQ Sequence 52 AA;

Query Match 86.9%; Score 259; DB 23; Length 52;
Best Local Similarity 92.5%; Pred. No. 1.5e-13;
Matches 49; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAKVPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52
DB 1 GAKAPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 52

```


Sequence 911 AA;

PI Charles IG;

Polypeptides containing polymorphisms of the repeated regions of pertactin in *Bordetella* species, useful in immunogenic compositions for treating infections caused by *Bordetella* and in diagnostic methods -

Search completed: May 7, 2003, 16:47:10
Job time : 28.0187 secs


```
RP SEQUENCE FROM N.A.
RC STRAIN-16039;
RA Kell D.J., Fenwick B.;
RT "Variability in the pertactin genes of Bordetella bronchiseptica
RT isolates includes regions coding for the GGXP and PGP families of
RT amino acid repeats";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156773; AAF82397.1; -.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 122 1
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 12395 MW; 8ED00966A40FF994 CRC64;

Query Match          94.9%; Score 258; DB 2; Length 122;
Best Local Similarity 85.7%; Pred. No. 9.5e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 GAKAPPAPKPA-----PQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
    |||||
Db 42 GAKAPPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 97

RESULT 14
Q9ALQ2 PRELIMINARY; PRT; 252 AA.
AC Q9ALQ2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pertactin (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC13935;
RX MEDLINE=21117018; PubMed=11179374;
RA Register K.B.;
RT "Novel Genetic and Phenotypic Heterogeneity in Bordetella
RT bronchiseptica pertactin.";
RL Infect. Immun. 69:1917-1921(2001).
DR EMBL; AY007265; AAC38441.1; -.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR004899; Pertactin_sup.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 26107 MW; 368C142508D77057 CRC64;

Query Match          94.9%; Score 258; DB 2; Length 252;
Best Local Similarity 85.7%; Pred. No. 1.7e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 GAKAPPAPKPA-----PQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
    |||||
Db 170 GAKAPPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 225

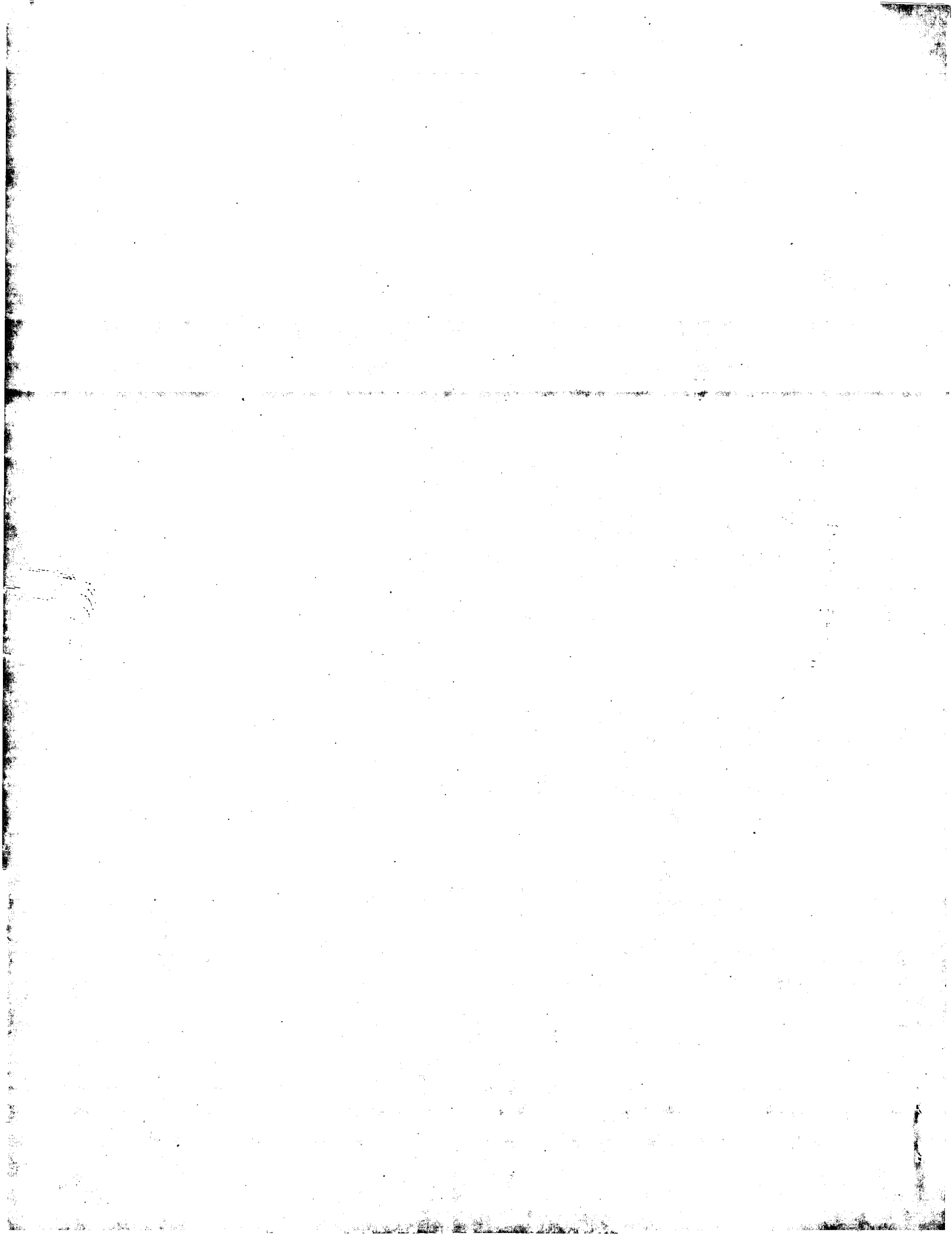
RESULT 15
Q9K5G7 PRELIMINARY; PRT; 111 AA.
ID Q9K5G7
AC Q9K5G7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE Pertactin (P.68) (Fragment).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DEL;
RX MEDLINE=20359389; PubMed=10899896;
RA Boursaux-Ende C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella
RT pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250089; CAB76443.1; -.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 11453 MW; 9B67012D3B9AECA CRC64;

Query Match          94.5%; Score 257; DB 2; Length 111;
Best Local Similarity 82.8%; Pred. No. 1.1e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

Qy 1 GAKAPPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
    |||||
Db 19 GAKAPPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 76

Search completed: May 7, 2003, 16:53:15
Job time : 24.5315 secs
```



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 4.66099 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754b-20

Perfect score: 272

Sequence: 1 GAKAPPAPKAPQCPQPPQ.....QRQPEAPAPQPPAGRELSAA 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	260	95.6	911	1 PERT_BORBR	Q03035 bordetella
2	257	94.5	922	1 PERT_BORPA	P24328 bordetella
3	223	82.0	910	1 PERT_BORPE	P14283 bordetella
4	127	46.7	485	1 SSGP_VOLCA	P21997 volvox cart
5	127	46.7	487	1 EBN2_EBV	P12978 epstein-bar
6	124	45.6	141	1 YPRO_OWEFU	P21260 owenia fusi
7	123	45.2	1362	1 BRD4_HUMAN	O60885 homo sapien
8	122.5	45.0	449	1 APC_BRANA	P40503 brassica na
9	120.5	44.3	2004	1 MOZ_HUMAN	Q92794 homo sapien
10	119.5	43.9	3110	1 HD_RAT	P51111 rattus norv
11	119	43.8	440	1 G3PT_MOUSE	O64467 mus musculu
12	118.5	43.6	808	1 Y066_NPVAC	P4467 autographa
13	118.5	43.6	1139	1 KPCL_TRIRE	Q99014 trichoderma
14	118	43.4	875	1 Y066_NPVOP	O83949 orgyia pseu
15	117.5	43.2	426	1 EXLP_TOBAC	Q03211 nicotiana t
16	117.5	43.2	443	1 HXA3_MOUSE	P02831 mus musculu
17	117.5	43.2	555	1 GPI_CHLRE	Q98qg6 chlamydomon
18	117	43.0	514	1 MEFD_MOUSE	Q65943 mus musculu
19	116	42.6	3144	1 HD_HUMAN	P42858 homo sapien
20	115	42.3	431	1 ACRO_RABIT	P48038 oryctolagus
21	115	42.3	534	1 APC_ARATH	P40502 arabidopsis
22	114	41.9	2715	1 TRX2_HUMAN	Q9um66 homo sapien
23	113	41.5	3119	1 HD_MOUSE	P42859 mus musculu
24	112.5	41.4	1790	1 SEPA_EMENT	P78621 emericeila
25	112.5	41.4	3164	1 TEGU_HSV11	P10220 herpes slimp
26	109.5	40.3	367	1 CSP_PLAYO	P06914 plasmodium
27	109	40.1	415	1 ACRO_PIG	P08001 sus scrofa
28	108	39.7	267	1 EXTN_MAIZE	P14918 zea mays (m
29	108	39.7	979	1 RFX1_HUMAN	P22670 homo sapien
30	107.5	39.5	232	1 ACCL_HUMAN	P58840 homo sapien
31	107.5	38.5	251	1 HXB4_HUMAN	P17483 homo sapien
32	107.5	37.9	379	1 YP85_CAEEL	Q09442 caenorhabdi
33	107	39.3	815	1 MK07_HUMAN	Q13164 homo sapien

RESULT 1

ID	PERT_BORBR	STANDARD;	PRT;	911 AA.
AC	Q03035;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pertactin precursor (Outer membrane protein P.68) (P.94).			
GN	PRN.			
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=518;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.			
RC	STRAIN=CN7531;			
RX	MEDLINE=92407514; PubMed=1527510;			
RA	Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;			
RT	"Cloning, nucleotide sequence and heterologous expression of the protective outer-membrane protein P.68 pertactin from Bordetella bronchiseptica."			
RL	J. Gen. Microbiol. 138:1697-1705(1992).			
CC	-!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.			
CC	-!- SUBUNIT: MONOMER.			
CC	-!- SUBCELLULAR LOCATION: Outer membrane.			
CC	-!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.			
CC	-!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+) CONCENTRATIONS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X54815; CAA38584.1; -			
DR	EMBL; A19180; CAA01453.1; -			
DR	PIR; A47675; A47675.			
DR	InterPro; IPR004899; Pertactin_sup.			
DR	InterPro; IPR003992; pertactin.			
DR	InterPro; IPR003991; pertactin_vir.			
DR	Pfam; PF03212; Pertactin; 1.			
DR	PRINTS; PR01482; PERTACTIN.			
DR	PRINTS; PR01484; PERTACTINFAMLY.			
DR	Outer membrane; Signal; Virulence; Repeat.			
KW	OUTER MEMBRANE; SIGNAL; VIRULENCE; REPEAT.			
FT	SIGNAL 1 34			
FT	CHAIN 35 911			
FT	CHAIN 35 712			
FT	PROPEP 713 911			
FT	SITE 260 262			
FT	-----			
FT	P.94.			
FT	PERTACTIN (P.68).			
FT	POTENTIAL.			
FT	CELL ATTACHMENT SITE (INVOLVED IN			
FT	ADHESION TO VARIOUS EUKARYOTIC CELL			
FT	LINES).			

Q24120 drosophila
P10323 homo sapien
O43365 homo sapien
O62203 mus musculu
Q14003 homo sapien
P05142 mus musculu
O06432 neisseria g
P05143 mus musculu
Q15428 homo sapien
Q05859 mus musculu
Q05860 mus musculu
P06915 plasmodium

ALIGNMENTS

```

FT SITE 701 703 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 266 275 3 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3 (APPROXIMATE).
FT DOMAIN 570 601 7 X 3 AA REPEATS OF P-Q-P.
FT SEQUENCE 911 AA; 93995 MW; 3078DF6EC2D987A1 CRC64;
Query Match 95.6%; Score 260; DB 1; Length 911;
Best Local Similarity 92.3%; Pred. No. 3.1e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPPAPKPA---PQCPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
Db 559 GAKAPPAPKPA---PQCPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 610

RESULT 2
PERT_BORPA
ID PERT_BORPA STANDARD; PRT; 922 AA.
AC P24328;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.70) (P.95).
CN PRN.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CN2591;
RX MEDLINE=91251771; PubMed=2041476;
RA Li L.J., Dougan G., Novotny P., Charles I.G.;
RT "P.70 pertactin, an outer-membrane protein from Bordetella
RT parapertussis: cloning, nucleotide sequence and surface expression in
RT Escherichia coli."
RL Mol. Microbiol. 5:409-417(1991).
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54547; CAA38419.1; -
DR EMBL; A26124; CAA01786.1; -
DR EMBL; A19182; CAA01454.1; -
DR PIR; S15204; S15204.
DR PIR; S14659; S14659.
DR InterPro; IPR004899; Pertactin_sup.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_vir.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PERTACTINFAMILY.
DR Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 922 P.95.
FT CHAIN 35 647 PERTACTIN (P.70).
FT PROPEP 648 922 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN

```

```

FT SITE 266 290 ADHESION TO VARIOUS EUKARYOTIC CELL
FT DOMAIN 266 270 LINES).
FT REPEAT 271 275 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 276 280 1.
FT REPEAT 281 285 2.
FT REPEAT 286 290 3.
FT DOMAIN 575 603 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
FT SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
Query Match 94.5%; Score 257; DB 1; Length 922;
Best Local Similarity 82.8%; Pred. No. 4.8e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 GAKAPPAPKPA---PQCPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
Db 564 GAKAPPAPKPA---PQCPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 621

RESULT 3
PERT_BORPE
ID PERT_BORPE STANDARD; PRT; 910 AA.
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
CN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CN2992;
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis."
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RN [2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica."
RL J. Gen. Microbiol. 138:1697-1705(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RA Emaley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin."
RL Nature 381:90-92(1996).
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04560; AAA22980.1; ALT_SEQ.

```

```

FT DOMAIN      228   340          PRO-RICH.
FT DOMAIN      260   295          POLY-PRO.
SQ SEQUENCE    485 AA;  50436 MW;  A5221640A0A031421 CRC64;

Query Match              46.7%; Score 127; DB 1; Length 485;
Best Local Similarity    54.1%; Pred. No. 0.038;
Matches                  20; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY      4 APPAPKAPDGPDPDPDPDPDPDPDPDPDPDPDPDPDPDP 40
DB      252 SPPPSPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 288
       :|||:|||:|||:|||:|||:|||:|||
Query Match              46.7%; Score 127; DB 1; Length 485;
Best Local Similarity    54.1%; Pred. No. 0.038;
Matches                  20; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

RESULT 5
EBN2_EBV      EBN2_EBV      STANDARD; PRT; 487 AA.
ID            AC           PI2978;
DT DT      01-OCT-1989 (Rel. 12, Created)
DT DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT DT      01-NOV-1995 (Rel. 32, Last annotation update)
DE DE      EBNA-2 nuclear protein.
GN GN      BVRF1.
OS OS      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC OC      Gammaherpesvirinae; Lymphocryptovirus.
NC NC      CBI_TaxID=10377;
RN RN      [1]
RN RN      SEQUENCE FROM N.A.
RX RX      MEDLINE=84270667; PubMed=6087149;
RA RA      Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA RA      Tuffnell P.S., Barrell B.G.;
RT RT      "DNA sequences and expression of the B95-8 Epstein-Barr virus genome.";
RL RL      Nature 310:207-211(1984).
RN RN      [2]
RX RX      SUBCELLULAR LOCATION, AND PHOSPHORYLATION.
RP RP      MEDLINE=90285473; PubMed=2161150;
RA RA      Pettit L., Sample C., Kieff E.;
RT RT      "Subnuclear localization and phosphorylation of Epstein-Barr virus
RL RL      latent infection nuclear proteins.";
RL RL      Virology 176:563-574(1990).
RN RN      [3]
RP RP      DOMAINS.
RX RX      MEDLINE=91202599; PubMed=18500028;
RA RA      Cohen J.I., Wang F., Kieff E.;
RT RT      "Epstein-Barr virus nuclear protein 2 mutations define essential
RL RL      domains for transformation and transactivation.";
RL RL      J. Virol. 65:2545-2554(1991).
CC CC      -1- FUNCTION: INVOLVED IN LATENT CYCLE. TRANSACTIVATES THE EXPRESSION
CC CC      OF LMP-1.
CC CC      -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC CC      -1- PTM: PHOSPHORYLATED.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement
or send an email to licensee@isb-sib.ch).
-----
EMBL; IVO1555; CAA24877.1; ALT_INIT.
DR DR      TRANSFAC; TO1618;
DR DR      Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW KW      Phosphorylation; Repeat.
FT FT      DOMAIN 345 356          POLY-PRO.
FT FT      SEQUENCE 487 AA; 52544 MW; DEF40DF8ED61D1A CRC64;
SQ SQ      6 X 2 AA TANDEM REPEATS OF R-G.

```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	260	95.6	911	2	A47675	68K outer membrane	
2	257	94.5	922	2	S15204	pertactin - Bordet	
3	223	82.0	930	2	A32560	outer membrane pro	
4	138.5	50.9	430	2	JC2301	hypothetical 47.8K	
5	128	47.1	599	2	T10798	phosphorin-S - VO	
6	127	46.7	485	2	A33647	sulfated surface g	
7	127	46.7	487	2	S42442	nuclear protein EB	
8	127	46.7	929	2	C96623	hypothetical prote	
9	126	46.3	1684	2	T03632	hypothetical prote	
10	124	45.6	141	2	A34043	hypothetical proli	
11	122.5	45.0	449	2	S16748	proline-rich prot	
12	121.5	44.7	446	2	T07907	hydroxyproline-ric	
13	121	44.5	1952	2	T48814	hypothetical prote	
14	120	44.1	209	2	A48232	cysteine-rich exte	
15	120	44.1	534	2	S21961	proline-rich prote	
16	119.5	43.9	544	2	T17547	proline-rich prote	
17	119	43.8	440	2	A3681	glyceraldehyde-3-p	
18	118.5	43.6	808	2	C72858	AcOrf-66 protein -	
19	118.5	43.6	1139	1	S61918	protein kinase C (
20	118.5	43.6	1494	2	T14355	protein-tyrosine-p	
21	118	43.4	383	2	AE2295	hypothetical prote	
22	118	43.4	875	2	T10340	hypothetical prote	
23	117.5	43.2	426	2	J01696	piatit extensin-li	
24	117.5	43.2	439	2	S51939	chitinase (EC 3.2.	
25	117	43.0	506	2	B56201	transcription fact	
26	117	43.0	514	2	A56201	transcription fact	
27	116.5	42.8	1388	2	T00063	hypothetical prote	
28	116	42.6	3144	2	A46068	Huntington disease	
29	115.5	42.5	1474	2	T18281	hypothetical prote	

Qy 1 GAKAPPAKPAQPG-----PQPPPPPPPPPPQPPQPPQPPQPPQPPQPPAGRELSAA 48

RESULT 8

```

A:Accession: A34043
A:Molecule type: DNA
A:Residues: 1-141 <BAK>
A:Cross-references: GB:M32217
A:Accession: B34043
A:Molecule type: DNA
A:Residues: 59-136 <BA2>
A:Cross-references: GB:M32217

```

```
Query Match      45.6%; Score 124; DB 2; Length 141;
Best Local Similarity 53.8%; Pred. No. 0.019;
Matches 21; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
```

A;Molecule type: DNA

proline-rich protein - rape (fragment)
 C/Species: Brassica napus (rape)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
 C/Accession: S16748

R; Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, J.; Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, J.

A;Description: Cloning and characterization of a proline-rich gene expressed sp
A;Reference number: S16748

A;Accession: S10748
A;Molecule type: mRNA
A;Residues: 1-449 <ROB>
A;Cross-references: EMBL:X60376; NID:q225996; PID:q22597

[illegible]

11 KPMFPPFAITFSCFCFPPFAPQPPAPPTFSCFCFPPFAPQPPAPGSPKPG 0 /

RESULT 12

T07907

hydroxyproline-rich glycoprotein GAS28 precursor - Chlamydomonas reinhardtii
C; Species: Chlamydomonas reinhardtii
C; Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C; Accession: T07907
R; Rodriguez-Martinez, H.; Haring, M.A.; Von Gromof, E.; Beck, C.F.
submitted to the EMBL Data Library, July 1997

A;Reference number: Z16207
A;Accession: T07907
A;Status: preliminary; translated from GB/EMBL/DDRI

A: molecule type: mRNA
A: Residues: 1-446 <ROD>
A: Cross-references: EMBL:AF015883; NID:G2384727; PIDN:AAB69862.1; PID:G2384738
A: Experimental source: gametes

F: 1-29/Domain: signal sequence #status predicted <SIG>
P: 30-446/Product: hydroxyproline-rich glycoprotein <HPRG>
C: Genetics:[]
A: Gene: GAS28 #status predicted <GAS28>

Query Match	44.7%;	Score 121.5;	DB 2;	Length 446;
Best Local Similarity	51.1%;	Pred. No. 0.07;		
Matches 23:	Conservative	4:	Mismatches	11:
	Indels	7:		

Qy 5 PPAPKPAQPGPQP-----PQPPQP-PQRQPEAPAPQPPA 41
|||:||| ||| ||| ||| : | :| |||
Db 85 PPRPRPCSPPPPPRRRCPSPPPPQPPRRRSPSPSPSPPPPA 127

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-14

Query Match 96.1%; Score 261.5; DB 9; Length 49;
Best Local Similarity 98.0%; Pred. No. 1.6e-12;
Matches 48; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAKAPPAPKPAQPGQP-PQPGPQPQPQPQPQPQPQPQPAPAPAPAGRELSAA 48
Db 1 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 49

RESULT 3
US-09-855-754-15
; Sequence 15, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-15

Query Match 95.6%; Score 260; DB 9; Length 52;
Best Local Similarity 92.3%; Pred. No. 2.2e-12;
Matches 48; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPPAPKPA-----PQPGPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 48
Db 1 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 52

RESULT 4
US-09-855-754-4
; Sequence 4, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-4

Query Match 95.6%; Score 260; DB 9; Length 911;
Best Local Similarity 92.3%; Pred. No. 2.3e-11;
Matches 48; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPPAPKPA-----PQPGPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 48
Db 559 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 610

RESULT 5
US-09-855-754-18
; Sequence 18, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bordetella bronchiseptica
US-09-855-754-18

Query Match 94.9%; Score 258; DB 9; Length 56;
Best Local Similarity 85.7%; Pred. No. 3.2e-12;
Matches 48; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GAKAPPAPKPA-----PQPGPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 48
Db 1 GAKAPPAPKPAQPGQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPAGRELSAA 56

RESULT 6
US-09-855-754-19
; Sequence 19, Application US/09855754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAX- EUDE, CAROLINE
; APPLICANT: GUISSO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855,754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 19

; LENGTH: 58

; TYPE: PRT

; ORGANISM: Bordetella bronchiseptica

US-09-855-754-19

Query Match 94.5%; Score 257; DB 9; Length 58;

Best Local Similarity 82.8%; Pred. No. 3.8e-12;

Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 GAKAPPAPKAPQPG-----PQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 48

Db 1 GAKAPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 58

RESULT 7

US-09-855-754-6

; Sequence 6, Application US/09855754

; Publication No. US20020192237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX- EUDE, CAROLINE

; APPLICANT: GUIISO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754

; PRIOR FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/206,969

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 922

; TYPE: PRT

; ORGANISM: Bordetella parapertussis

US-09-855-754-6

Query Match 94.5%; Score 257; DB 9; Length 922;

Best Local Similarity 82.8%; Pred. No. 3.8e-11;

Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 GAKAPPAPKAPQPG-----PQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 48

Db 564 GAKAPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 621

RESULT 8

US-09-855-754-21

; Sequence 21, Application US/09855754

; Publication No. US20020192237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX- EUDE, CAROLINE

; APPLICANT: GUIISO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754

; PRIOR FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/206,969

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 52

; TYPE: PRT

; ORGANISM: Bordetella bronchiseptica

US-09-855-754-21

Query Match 91.5%; Score 249; DB 9; Length 52;

Best Local Similarity 88.5%; Pred. No. 1.3e-11;

Matches 46; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GAKAPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 48

Db 1 GAKVPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 52

RESULT 9

US-09-855-754-22

; Sequence 22, Application US/09855754

; Publication No. US20020192237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX- EUDE, CAROLINE

; APPLICANT: GUIISO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754

; CURRENT FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/206,969

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 22

; LENGTH: 54

; TYPE: PRT

; ORGANISM: Bordetella bronchiseptica

US-09-855-754-22

Query Match 91.2%; Score 248; DB 9; Length 54;

Best Local Similarity 85.2%; Pred. No. 1.5e-11;

Matches 46; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

QY 1 GAKAPPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 48

Db 1 GAKVPAPKAPQPGQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAPAPQPPAGRELSAA 54

RESULT 10

US-09-855-754-17

; Sequence 17, Application US/09855754

; Publication No. US20020192237A1

; GENERAL INFORMATION:

; APPLICANT: BOURSAX- EUDE, CAROLINE

; APPLICANT: GUIISO-MACLOUF, NICOLE

; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED

; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,

; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA

; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS

; FILE REFERENCE: 03495-0206-00000

; CURRENT APPLICATION NUMBER: US/09/855,754

; CURRENT FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/206,969

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 17

; LENGTH: 52

; TYPE: PRT

; ORGANISM: Bordetella bronchiseptica

US-09-855-754-17

Query Match 89.7%; Score 244; DB 9; Length 52;

Best Local Similarity 88.7%; Pred. No. 2.8e-11;

Matches 47; Conservative 0; Mismatches 0; Indels 6; Gaps 2;

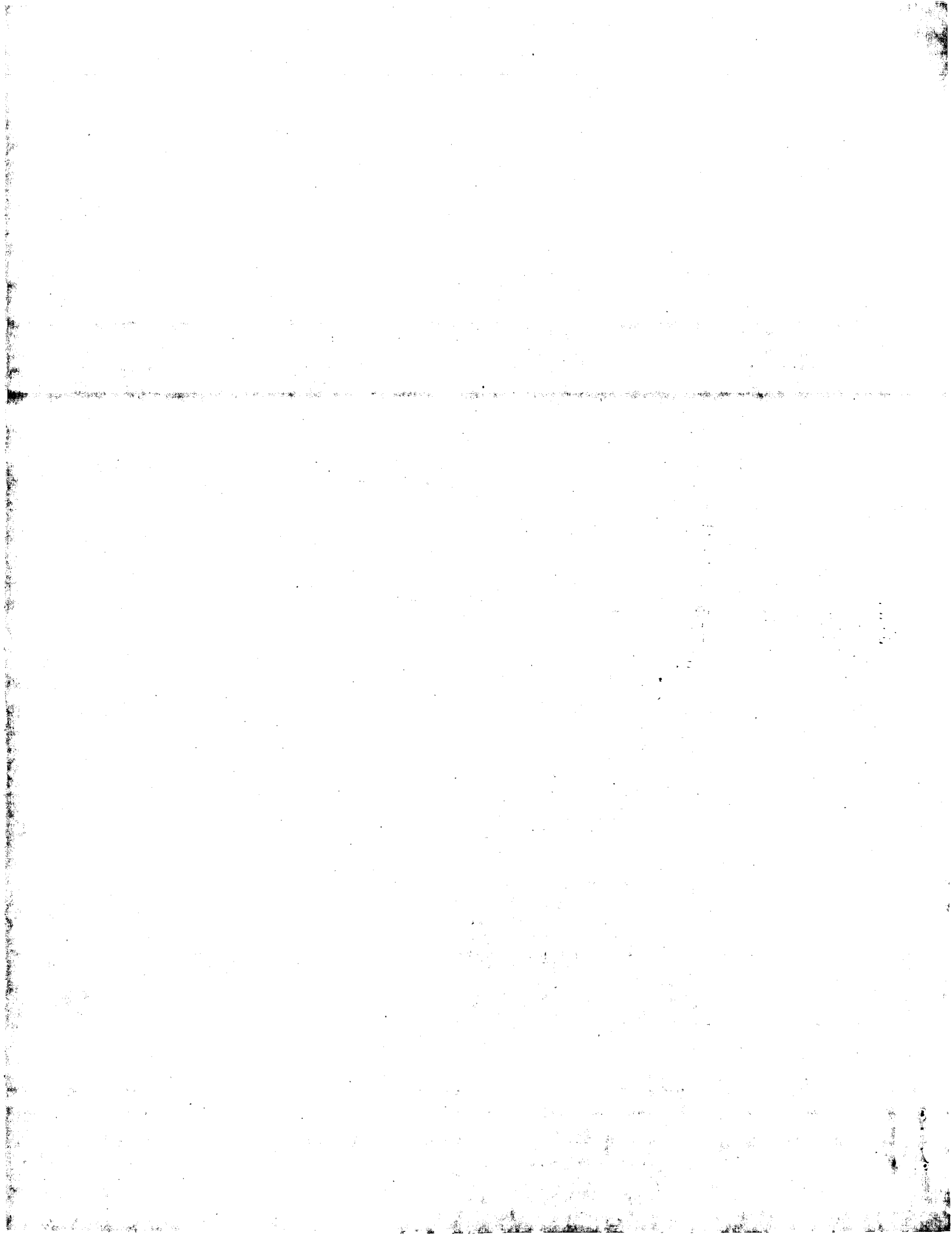
RESULT 15
US-09-976-740-43
; Sequence 43, Application US/09976740
; Publication No. US20020194633A1

GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-43

Query Match 43.2%; Score 117.5; DB 9; Length 538;
Best Local Similarity 56.5%; Pred. No. 0.13;
Matches 26; Conservative 2; Mismatches 11; Indels 7; Gaps 3;

QY 2 AKAPPA-PKPAPQPG-----PQPPQPPQPPQPP-QRQPEAPAPQPP 40
Db 189 AVAPPAGPRRAPPAVAAREPLPPPPQPPAPPQQQQPPPPQPP 234

Search completed: May 7, 2003, 17:30:54
Job time: 15.3918 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:38:52 ; Search time 29.6422 Seconds
(without alignments)
403.166 Million cell updates/sec

Title: US-09-855-754B-19
Perfect score: 335
Sequence: 1 GAKAPPAPKAPQPGQPGP.....QKQPEAPAPQPPAGRELSAA 58

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	335	100.0	111	2	Q9K5G7
2	314	93.7	115	2	Q9K5H3
3	303	90.4	111	2	Q9K5G9
4	303	90.4	122	2	Q9K5X9
5	303	90.4	252	2	Q9ALQ2
6	284	84.8	107	2	Q9K5H5
7	284	84.8	200	2	Q9ALP9
8	284	84.8	911	2	Q9L4E2
9	282.5	84.3	109	2	Q9K5H1
10	282.5	84.3	215	2	Q9ALQ4
11	265	79.1	107	2	Q9K5G1
12	263.5	78.7	104	2	Q9K5H9
13	263.5	78.7	105	2	Q9K5H7
14	263.5	78.7	115	2	Q9K5Y1
15	263.5	78.7	115	2	Q9K5Y0
16	263.5	78.7	198	2	Q9AHP0

17	257	76.7	105	2	Q9K5G5	Q9k5g5 bordetella
18	256	76.4	107	2	Q9K5G4	Q9k5g4 bordetella
19	249	74.3	231	2	Q9ALP3	Q9alp3 bordetella
20	244.5	73.0	304	2	Q9ALP7	Q9alp7 bordetella
21	231	69.0	256	2	Q9ALQ0	Q9alq0 bordetella
22	224	66.9	347	2	Q9ALP5	Q9alp5 bordetella
23	209	62.4	851	2	Q9AIX8	Q9aix8 bordetella
24	209	62.4	905	2	Q9SGM9	Q9sgm9 bordetella
25	209	62.4	905	2	Q9S3M8	Q9s3m8 bordetella
26	209	62.4	905	2	Q9SGN0	Q9sgn0 bordetella
27	209	62.4	910	2	Q9SGN1	Q9sgn1 bordetella
28	209	62.4	910	2	Q9G2S9	Q9g2s9 bordetella
29	209	62.4	915	2	Q9G2S9	Q9g2s9 bordetella
30	209	62.4	920	2	Q9S1P8	Q9s1p8 bordetella
31	188.5	56.3	907	2	Q9G2S7	Q9g2s7 bordetella
32	188.5	56.3	912	2	Q9RSU0	Q9rsu0 bordetella
33	168.5	50.3	430	3	Q01823	Q01823 pneumocysti
34	167.5	50.0	1217	4	Q9ULL5	Q9ull5 homo sapien
35	167	49.9	938	3	Q9GVI4	Q9gvi4 pneumocysti
36	165	49.3	599	10	P93797	P93797 volvox cart
37	163	48.7	243	5	Q8TIY6	Q8tiy6 dictyostell
38	162	48.4	409	10	Q9SBM1	Q9sbm1 volvox cart
39	162	48.4	1143	10	Q948Y6	Q948y6 volvox cart
40	159.5	47.6	493	3	Q9UVD1	Q9uud1 pneumocysti
41	159	47.5	203	16	Q9RCX9	Q9rcx9 streptomyc
42	157.5	47.0	1011	3	Q9P944	Q9p944 pneumocysti
43	157	46.9	166	3	Q9UVD0	Q9uud0 pneumocysti
44	156	46.6	383	16	Q8YQB7	Q8yqb7 anabaena sp
45	156	46.6	490	12	Q69023	Q69023 human herpe

ALIGNMENTS

RESULT 1	Q9K5G7	PRELIMINARY;	PRT;	111 AA.
ID	Q9K5G7	PRELIMINARY;	PRT;	111 AA.
AC	Q9K5G7	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 21, Last annotation update)			
DE	Pertactin (P.68) (Fragment).			
GN	PRN.			
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).			
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;			
OC	Bordetella.			
OX	NCBI_TaxID=518;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-DEL:			
RX	MEDLINE=20359389; PubMed=10899896;			
RA	Boursaux-Eude C., Guiso N.;			
RT	"Polymorphism of Repeated Regions of Pertactin in Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";			
RL	Infect. Immun. 68:4815-4817(2000).			
DR	EMBL: AJ250089; CAB76443.1;			
DR	InterPro: IPR004899; Pertact_sup.			
DR	InterPro: IPR002965; P-rich_extensn.			
DR	Pfam: PF03212; Pertactin; 1.			
FT	PRINTS; PR01217; PRICEXTENS.			
FT	NON_TER 1			
FT	NON_TER 111			
SQ	SEQUENCE 111 AA; 11453 MW; 9B67012D3B9AECA CRC64;			

Query Match 100.0%; Score 335; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. NO. 3.3e-21;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAKAPPAPKAPQPGQPGQPGP.....QKQPEAPAPQPPAGRELSAA 58
Db	19	GAKAPPAPKAPQPGQPGQPGQPGP.....QKQPEAPAPQPPAGRELSAA 76

RX	MEDLINE-20359389; PubMed=10899896;
RT	Boursaux-Eude C., Gulso N.;
RA	"Polymorphism of Repeated Regions of Pertactin in Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica.";
RL	Infect. Immun. 68:4815-4817(2000).
DR	ENBL: AJ250085; CAB76439.1; -
DR	InterPro: IPR004899; Pertact_sup.
DR	Pfam: PF03212; Pertactin; 1.
FT	NON_TER 1
FT	NON_TER 109
SEQUENCE	109 AA; 11192 MW; 85EBEF748366DD3 CRC64;
Query Match	84.3%; Score 282.5; DB 2; Length 109;
Best Local Similarity	89.7%; Pred. No. 6.8e-17;
Matches	52; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
QY	1 GAKAPPAPKPAQGPGQPQQPQQPPQPPQPPQPPQPEAPAPQPPAGRELSAA 58 DB
DB	22 GAKAPPAPKPAQGPGGPQP-----GQPPQPPQPPQPEAPAPQPPAGRELSAA 74
RESULT 10	
Q9ALQ4	
ID	PRELIMINARY; PRT; 215 AA.
AC	Q9ALQ4;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Pertactin (Fragment).
GN	PRN.
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OX	Bordetella.
OX	NCBI_TaxID=518;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-SOI0328;
RX	MEDLINE-21117018; PubMed=11179374;
RA	Register K.B.;
RT	"Novel Genetic and Phenotypic Heterogeneity in Bordetella bronchiseptica Pertactin";
RL	Infect. Immun. 69:1917-1921(2001).
DR	ENBL: AY007263; AAG38439.1; -
DR	InterPro: IPR003992; pertactin.
DR	InterPro: IPR004899; Pertact_sup.
DR	InterPro: IPR002965; P_rich_extensn.
DR	Pfam: PF03212; Pertactin; 1.
DR	PRINTS: PR01482; PERTACTIN.
DR	PRINTS: PR01217; PRICHEXTENS.
FT	NON_TER 1
FT	NON_TER 215
SEQUENCE	215 AA; 23237 MW; 5C21D45CF784B4AE CRC64;
Query Match	84.3%; Score 282.5; DB 2; Length 215;
Best Local Similarity	89.7%; Pred. No. 1.2e-16;
Matches	52; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
QY	1 GAKAPPAPKPAQGPGQPQQPQQPQQPQQPQQPQQPEAPAPQPPAGRELSAA 58 DB
DB	140 GAKAPPAPKPAQGPGGPQP-----GQPPQPPQPPQPEAPAPQPPAGRELSAA 192
RESULT 11	
Q9K5G1	
ID	PRELIMINARY; PRT; 107 AA.
AC	Q9K5G1;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Pertactin (P.68) (Fragment).
GN	PRN.
OS	Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

```
GN Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OS Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LAPR;
RX MEDLINE=20359389; PubMed=10899896;
RA Boursaux-Eude C., Guiso N.;
RT "Polymorphism of Repeated Regions of Pertactin in Bordetella
RL Infect. Immun. 68:4815-4817(2000).
DR EMBL; AJ250079; CAB76433.1; -.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
FT NON_TER 1
FT NON_TER 105
SQ SEQUENCE 105 AA; 10861 MW; 1F97255EB792F4B7 CRC64;

Query Match 78.7%; Score 263.5; DB 2; Length 105;
Best Local Similarity 84.5%; Pred. No. 2.4e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPPAPKAPQPGQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 58
Db 22 GAKAPPAPKAPQPGQPG-----PQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 70

RESULT 14
Q9KJY1
ID Q9KJY1 PRELIMINARY; PRT; 115 AA.
AC Q9KJY1
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Pertactin (Fragment).
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RA Keil D.J., Fenwick B.;
RT "Variability in the pertactin genes of Bordetella bronchiseptica
RL Isolates includes regions coding for the GGXP and PGP families of
RT amino acid repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156771; AAF82395.1; -.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01217; PRICHTEXTENS.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 11693 MW; FA3A76596F7097EB CRC64;

Query Match 78.7%; Score 263.5; DB 2; Length 115;
Best Local Similarity 84.5%; Pred. No. 2.6e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPPAPKAPQPGQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 58
Db 42 GAKAPPAPKAPQPGQPG-----PQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 90

RESULT 15
Q9KJY0
ID Q9KJY0 PRELIMINARY; PRT; 115 AA.
AC Q9KJY0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pertactin (Fragment).
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B;
RA Keil D.J., Fenwick B.;
RT "Variability in the pertactin genes of Bordetella bronchiseptica
RL Isolates includes regions coding for the GGXP and PGP families of
RT amino acid repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156772; AAF82396.1; -.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF03212; Pertactin; 1.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 11705 MW; ELABD4D347D20652 CRC64;

Query Match 78.7%; Score 263.5; DB 2; Length 115;
Best Local Similarity 84.5%; Pred. No. 2.6e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 GAKAPPAPKAPQPGQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 58
Db 42 GAKAPPAPKAPQPGQPG-----PQPPQPPQPPQPPQPPQPPQPPQPPQPPAGRELSAA 90

Search completed: May 7, 2003, 16:53:15
Job time : 30.6422 secs
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:28 ; Search time 24.9404 Seconds
(without alignments)
256.453 Million cell updates/sec

Title: US-09-855-754B-20

Perfect score: 272

Sequence: 1 GAKAPPAPAPQPGPPQ.....QRQPEAPAPQAGRELSAA 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	272	100.0	48	AAE16199	B. bronchiseptica
2	261.5	96.1	49	AAE16193	B. bronchiseptica
3	260	95.6	52	AAE16194	B. bronchiseptica
4	260	95.6	911	AAE14320	Pertactin antigen
5	260	95.6	911	AAE26503	prn proteins. Bor
6	260	95.6	911	AAE16183	Bordetella bronchi
7	259.5	95.4	53	AAE16196	B. bronchiseptica
8	258	94.9	56	AAE16197	B. bronchiseptica
9	257	94.5	58	AAE16198	B. bronchiseptica
10	257	94.5	922	AAE25578	Bordetella paraper

11	257	94.5	922	23	AAE16185	Bordetella paraper
12	256	94.1	60	23	AAE16195	B. bronchiseptica
13	249	91.5	52	23	AAE16200	B. bronchiseptica
14	248	91.2	54	23	AAE16201	B. bronchiseptica
15	247	90.8	922	12	AAE14321	Pertactin antigen
16	223	82.0	42	23	AAE16202	B. bronchiseptica
17	223	82.0	910	23	AAE16184	Bordetella pertuss
18	223	82.0	910	23	AAE17146	Bordetella pertuss
19	202.5	74.4	39	23	AAE16203	B. bronchiseptica
20	165	60.7	31	12	AAE11739	Peptide BBO5 deriv
21	159	58.5	31	12	AAE11737	Peptide 683 derive
22	135.5	49.8	900	21	AAE42321	Human ORFX ORF2085
23	135.5	49.8	1217	22	ABG09876	Novel human diagno
24	135.5	49.8	1239	22	ABG09877	Novel human diagno
25	132	48.5	44	22	AAO04929	Human polypeptide
26	127	46.7	24	17	AAE87649	Synthetic Bordetel
27	124	45.6	80	22	AAO04412	Human polypeptide
28	124	45.6	85	22	AAO02036	Human polypeptide
29	124	45.6	99	22	AAO02076	Human polypeptide
30	123	45.2	85	22	AAO04573	Human polypeptide
31	123	45.2	88	22	AAO02124	Human polypeptide
32	122	44.9	446	22	ABG70063	Drosophila melanog
33	120.5	44.3	325	22	ABG21919	Novel human diagno
34	120	44.1	356	22	ABG04360	Novel human diagno
35	120	44.1	406	22	ABG27350	Novel human diagno
36	117.5	43.2	439	13	AAE28150	Sugar beet chitina
37	117.5	43.2	538	22	AAE82806	Human low density
38	117	43.0	125	22	ABG17580	Novel human diagno
39	116.5	42.8	350	23	ABG93838	Herbicideally activ
40	116.5	42.8	3119	18	AAW36888	Mouse huntingtin's
41	116.5	42.8	3119	19	AAW44743	Mouse huntingtin p
42	116	42.6	171	20	AAW99022	Human huntingtin p
43	116	42.6	513	20	AAW33500	Human huntingtin p
44	116	42.6	530	20	AAW33501	Human apopain clea
45	116	42.6	550	22	AAE82807	Rabbit low density

ALIGNMENTS

RESULT 1

AAE16199

ID AAE16199 standard; peptide; 48 AA.

XX AAE16199;

AC AAE16199;

XX 26-MAR-2002 (first entry)

DT 26-MAR-2002 (first entry)

XX B. bronchiseptica strain II-7 pertactin outer membrane protein region II.

DE Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;

KW therapy; antibiotic; antibacterial; region II.

XX Bordetella bronchiseptica.

OS Bordetella bronchiseptica.

XX WO200190143-A2.

PN 29-NOV-2001.

XX 23-MAY-2001; 2001WO-EP06457.

XX 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of

PT pertactin in Bordetella species, useful in immunogenic compositions for

PT treating infections caused by Bordetella and in diagnostic methods

XX

RESULT 4

AA14320
ID AAR14320 standard; Protein; 911 AA.

XX AC AAR14320;
XX DT 20-JAN-1992 (first entry)
XX DE Pertactin antigen P.68.
XX DE Pertactin; Pichia; B. pertussis; B. paraptussis.
XX KW Bordetella bronchiseptica.
XX OS

XX FH Key Location/Qualifiers
FT Peptide 266..270
FT Peptide /label= repeat
FT Peptide 271..275
FT Peptide /label= repeat
FT Peptide 570..572
FT Peptide /label= repeat
FT Peptide 574..576
FT Peptide /label= repeat
FT Peptide 578..580
FT Peptide /label= repeat
FT Peptide 581..583
FT Peptide /label= repeat
FT Peptide 584..586
FT Peptide /label= repeat
FT Peptide 587..589
FT Peptide /label= repeat
FT Peptide 599..601
FT Peptide /label= repeat

XX PN W09115571-A.
XX PD 17-OCT-1991.
XX PF 28-MAR-1991; 91WO-GB00487.
XX PR 02-APR-1990; 90GB-0007416.
XX PA (WELL) WELLCOME FOUNDATION LTD.
XX PI Clare JJ, Romanos MA;
XX PS WPI; 1991-325214/44.
XX DR N-PSDB; AAQ14319.
XX CC Pichia microorganism transformants - for production of
XX PT Bordetella pertactin antigens for whooping cough vaccines
XX PS Disclosure; Fig 1b; 38pp; English.
XX CC Pichia microorganisms are transformed for the expression of
XX CC pertactin antigens. DNA sequence used are represented in AAQ14319-20
XX CC encoding the B. bronchiseptica P.68 and B. paraptussis P.70 antigen
XX CC respectively or the B. pertussis P.69 encoding sequence described
XX CC by i.g. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
XX CC (1989).
XX SQ Sequence 911 AA;
Query Match 95.6%; Score 260; DB 12; Length 911;
Best Local Similarity 92.3%; Pred. No. 4e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPAPKPA-----PQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
|||||
DB 559 GAKAPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 610

RESULT 5

AAR26503
ID AAR26503 standard; Protein; 911 AA.
XX AC AAR26503;
XX DT 12-MAR-1993 (first entry)
XX DE prn proteins;
XX DE B. bronchiseptica; P.68; outer membrane protein; piglet; probe;
XX KW atrophic rhinitis; alternative cleavage.
XX OS

XX FH Key Location/Qualifiers
FT Protein 35..632
FT Peptide /label= P.68
FT Region 266..279
FT Region /label= Repeat_region
FT Region 570..589
FT Peptide 260..262
FT Peptide /label= RGD-tripeptide
FT Peptide 701..703
FT Peptide /label= RGD-tripeptide

XX PN W09217587-A.
XX PD 15-OCT-1992.
XX PF 27-MAR-1992; 92WO-GB00561.
XX PR 27-MAR-1991; 91GB-0006568.
XX PA (WELL) WELLCOME FOUND LTD.
XX PI Charles IG;
XX DR WPI; 1992-366258/44.
XX DR N-PSDB; AAQ34566.
XX CC DNA encoding a Bordetella bronchiseptica protein - used for
XX PT obtaining vaccines for preventing respiratory diseases, partic.
XX PT atrophic rhinitis in pigs
XX PS Claim 1; Fig 1; 28pp; English.
XX CC The sequence given is the P.94 antigen from B. bronchiseptica. The
XX CC P.68 antigen is formed by alternative cleavage of this protein.
XX CC P.68 is an outer membrane protein with a molecular weight of 68 kD
XX CC which is associated with protection of piglets against atrophic
XX CC rhinitis. The DNA sequence encoding these proteins was derived by
XX CC standard recombinant DNA techniques using P.68 probes to isolate the
XX CC entire P.94 sequence.
XX SQ Sequence 911 AA;
Query Match 95.6%; Score 260; DB 13; Length 911;
Best Local Similarity 92.3%; Pred. No. 4e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 GAKAPAPKPA-----PQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48
|||||
DB 559 GAKAPAPKPAQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 610

RESULT 6

AAE16183
ID AAE16183 standard; Protein; 911 AA.
XX AC AAE16183;
XX

PI Guiso-maclouf N, Boursaux-eude C;

DR WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

XX Claim 26; Fig 1c; 47pp; English.

XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections
CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region II.

XX Sequence 56 AA;

Query Match 94.9%; Score 258; DB 23; Length 56;

Best Local Similarity 85.7%; Pred. No. 7.1e-14;

Matches 48; Conservative 0; Mismatches 8; Indels 8; Gaps 1;

OY 1 GAKAPPAPKPA-----PQGPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPAGRELSAA 48

DB 1 GAKAPPAPKPAQPAGRELSAA 56

RESULT 9

AAE16198

ID AAE16198 standard; peptide; 58 AA.

XX AAE16198;

XX 26-MAR-2002 (first entry)

XX B. bronchiseptica strain II-6 pertactin outer membrane protein region II.

XX Pertactin; PRN; outer membrane protein; vaccine; Bordetella infection;

XX therapy; antibiotic; antibacterial; region II.

XX Bordetella bronchiseptica.

XX WO200190143-A2.

XX 29-NOV-2001.

XX 23-MAY-2001; 2001WO-EP06457.

XX 25-MAY-2000; 2000US-206969P.

XX (INSP) INST PASTEUR.

XX Guiso-maclouf N, Boursaux-eude C;

XX WPI; 2002-097639/13.

XX Polypeptides containing polymorphisms of the repeated regions of
PT pertactin in Bordetella species, useful in immunogenic compositions for
PT treating infections caused by Bordetella and in diagnostic methods

XX Claim 26; Fig 1c; 47pp; English.

XX The present invention relates to Bordetella bronchiseptica pertactin
CC (outer membrane protein) or their fragments. Pertactin (PRN) is used as
CC vaccine. Pertactin antibody is useful for treating Bordetella infections

CC and used to detect Bordetella antigens in biological preparations or in
CC purifying corresponding proteins, glycoproteins or their mixtures when
CC used in affinity chromatographic columns. Pertactin is useful as antigens
CC to identify antibodies to Bordetella in materials such as human or other
CC animal tissue and human or other animal cells, as well as biological
CC fluids, such as human or other animal body fluids, including human sera,
CC and to determine the concentration of Ab in those materials. Thus the
CC antigens can be used for qualitative or quantitative determination of
CC Bordetella in a material. The present sequence is B. bronchiseptica
CC pertactin outer membrane protein region II.

XX Sequence 53 AA;

Query Match 94.5%; Score 257; DB 23; Length 58;

Best Local Similarity 82.8%; Pred. No. 8.7e-14;

Matches 48; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY 1 GAKAPPAPKPAQPAGRELSAA 48

DB 1 GAKAPPAPKPAQPAGRELSAA 58

RESULT 10

AAR25578

ID AAR25578 standard; Protein; 922 AA.

XX AAR25578;

XX 08-JAN-1993 (first entry)

XX Bordetella parapertussis P95 antigen precursor.

XX Whooping cough; P70 antigen; P95 precursor protein; vaccination.

XX Bordetella parapertussis.

XX Key Location/Qualifiers

XX Protein 35..643

XX Binding-site /label= P70

XX 260..262

XX /note= "motif associated with cell-cell adhesion"

XX 266..285

XX /note= "contains 5 direct, tandem repeats"

XX 575..612

XX /note= "contains 9 direct repeats of Pro-Gln-Pro"

XX 712..714

XX /note= "motif associated with cell-cell adhesion"

XX WO9211292-A.

XX 09-JUL-1992.

XX 23-DEC-1991; 91WO-GB02302.

XX 21-DEC-1990; 90GB-0027901.

XX (WELL) WELLCOME FOUND LTD.

XX Charles IG;

XX WPI; 1992-250033/30.

XX N-PSDB; AAQ26509.

XX Acellular vaccine for immunisation against whooping cough -
PT comprises protein unadorned by B. parapertussis components
PT and capable of binding antibodies which bind native P70 antigen

XX Claim 1; Fig 1; 20pp; English.

XX A cosmid library was constructed by transforming E.coli HB101 with
CC recombinant cosmids prepared by partial digestion of B.parapertussis
CC chromosomal DNA with Sau3A and cloning of 40-50kb fragments into the
CC BamHI site of cosmid pHC79. The cosmids were screened with a 1.8kb

GAKAPPAQPGPQP

[illegible]

	QY	1 GAKAPPAIKPAAAA-PQPQQPPPPPQQRPQAAPPAGRELSAA 48
	D6	559 GAKAPPAIKPAPQGPGGPPPPPPPOROPEAPAOPPAGRELSAA 610

NAME/KEY: Region
LOCATION: 1..2
OTHER INFORMATION: /label= LTB sequence
FEATURE:
NAME/KEY: Region
LOCATION: 4..7
OTHER INFORMATION: /label= hinge region
FEATURE:
NAME/KEY: Region
LOCATION: 10..31
OTHER INFORMATION: /label= p69 BB05 epitope of Bordetella
OTHER INFORMATION: pertussis sequence
US-08-237-716-11

Query Match 48.9%; Score 133; DB 1; Length 33;
Best Local Similarity 70.6%; Pred. No. 4.7e-06;
Matches 24; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 6 PAKPAPQGPQPQPQPQPQPQPQPQPQP 39
DB 5 PGFEIAPQGPQPQPQPQPQPQPQPQP 32

RESULT 5
US-08-750-624-11
; Sequence 11, Application US/08750624
; Patent No. 6290971
; GENERAL INFORMATION:
; APPLICANT: Kandil, Ali
; APPLICANT: James, Olive A
; APPLICANT: Chong, Pele
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS COMPRISING A
; TITLE OF INVENTION: MINERAL SALT AND ANOTHER IMMUNOSTIMULATING COMPOUND
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-660
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Query Match 46.7%; Score 127; DB 4; Length 24;
Best Local Similarity 79.3%; Pred. No. 1.2e-05;
Matches 23; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 12 PQPGQPQPQPQPQPQPQPQPQPQPQP 40
DB 1 PQPGQPQPQPQPQPQPQPQPQPQP 23

RESULT 6
US-08-460-269C-8
; Sequence 8, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; APPLICANT: ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
; YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-460-269C-8

Query Match 45.6%; Score 124; DB 4; Length 23;
Best Local Similarity 79.3%; Pred. No. 2.1e-05;
Matches 23; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 11 APQGPQPQPQPQPQPQPQPQPQPQP 39
DB 1 APQGPQPQPQPQPQPQPQPQPQP 23

RESULT 7
US-08-246-982A-16
; Sequence 16, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: Macdonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1100 New York Avenue
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20005
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/246.982A
;; FILING DATE: May 20, 1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Goldstein, Jorge, A.
;; REGISTRATION NUMBER: 29,021
;; REFERENCE/DOCKET NUMBER: 0609.3880002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3144 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-08-246-982A-6

Query Match 42.6%; Score 116; DB 1; Length 3144;
Best Local Similarity 46.4%; Pred. No. 0.0085;
Matches 26; Conservative 1; Mismatches 15; Indels 14; Gaps 2;

QY 6 PAKPAPQPGQPQQPPQPQP--PQRQPEAPAPQPPAG-----RELSA 47
Db 41 PPPPPPPPPQLPQPQQPQAQLLPQPQPPPPPPPPGPAVAEPLHRPKKELSA 96

RESULT 14
US-08-453-265-6
; Sequence 6, Application US/08453265
; Patent No. 5693757
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,265
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3880003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-265-6

Query Match 42.6%; Score 116; DB 1; Length 3144;
Best Local Similarity 46.4%; Pred. No. 0.0085;
Matches 26; Conservative 1; Mismatches 15; Indels 14; Gaps 2;

QY 6 PAKPAPQPGQPQQPPQPQP--PQRQPEAPAPQPPAG-----RELSA 47
Db 41 PPPPPPPPPQLPQPQQPQAQLLPQPQPPPPPPPPGPAVAEPLHRPKKELSA 96

RESULT 15
US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-457-273B-42

Query Match 42.6%; Score 116; DB 2; Length 3144;
Best Local Similarity 46.4%; Pred. No. 0.0085;
Matches 26; Conservative 1; Mismatches 15; Indels 14; Gaps 2;

QY 6 PAKPAPQPGQPQQPPQPQP--PQRQPEAPAPQPPAG-----RELSA 47
Db 41 PPPPPPPPPQLPQPQQPQAQLLPQPQPPPPPPPPGPAVAEPLHRPKKELSA 96

Search completed: May 7, 2003, 16:57:24
Job time: 10.0228 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:31:49 ; Search time 5.63203 Seconds
(without alignments)
427.133 Million cell updates/sec

Title: US-09-855-754B-19

Perfect score: 335

Sequence: 1 GAKAPPAPKAPQGPQGPQ.....QRQEPAPQPPAGRELSAA 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	335	100.0	922	1 PERT_BORPA	P24328 bordetella
2	284	84.8	911	1 PERT_BORBR	Q03035 bordetella
3	209	62.4	910	1 PERT_BORPE	P14283 bordetella
4	166	49.6	485	1 SSGP_VOLCA	P21997 volvox cart
5	165.5	49.4	449	1 APG_BRANA	P40603 brassica na
6	161	48.1	141	1 IPKO_OMEFU	P21260 owenia fusi
7	152.5	45.5	487	1 EBN2_EBV	P12978 epstein-bar
8	150.5	44.9	431	1 ACRO_RABIT	P48038 cryptolagus
9	148	44.2	555	1 GPI_CHLRE	Q9fpg6 chlamydomon
10	148	44.2	875	1 Y066_NPVOP	Q83949 orgyia pseu
11	146.5	43.7	3164	1 TEGU_HSV11	P10220 herpes simp
12	141.5	42.2	534	1 APG_ARATH	P40602 arabidopsis
13	141	42.1	440	1 G3PT_MOUSE	Q64467 mus musculu
14	140.5	41.9	1206	1 FM14_MOUSE	Q05859 mus musculu
15	140.5	41.9	1468	1 FMN1_MOUSE	Q05860 mus musculu
16	139	41.5	2004	1 MOZ_HUMAN	Q92794 homo sapien
17	138	41.2	1790	1 SEPA_EMENI	P78621 emericeila
18	136.5	40.7	979	1 REX1_HUMAN	P22670 homo sapien
19	135.5	40.4	426	1 EXLP_TOBAC	Q03211 nicotiana t
20	135.5	40.4	1059	1 CAPU_DROME	P24120 drosophila
21	134.5	40.1	415	1 ACRO_PIG	P08001 sus scrofa
22	134	40.0	102	1 COLU_HSVSC	P22576 herpesvirus
23	133	39.7	464	1 S3A2_HUMAN	Q15428 homo sapien
24	132	39.4	475	1 S3A2_MOUSE	P25050 herpesvirus
25	131.5	39.3	105	1 COLL_HSVS7	P05142 mus musculu
26	131.5	39.3	261	1 PRP2_MOUSE	P05143 mus musculu
27	131.5	39.3	296	1 PRP3_MOUSE	Q99014 trichoderna
28	130.5	39.0	1139	1 KPC1_TIRE	P51111 rattus norv
29	130.5	39.0	3110	1 HD_RAT	Q08816 rattus norv
30	130	38.8	501	1 WASL_RAT	Q95107 bos taurus
31	130	38.8	505	1 WASL_BOVIN	Q00401 homo sapien
32	130	38.8	505	1 WASL_HUMAN	Q09442 caenorhabdi
33	129.5	38.7	379	1 YP85_CAEEL	

34 129 38.5 520 1 WASP_MOUSE
35 128.5 38.4 1362 1 BRD4_HUMAN
36 128 38.2 296 1 GDA6_WHEAT
37 127.5 38.1 443 1 HXA3_MOUSE
38 127.5 38.1 565 1 MOR8_MOUSE
39 126.5 37.8 129 1 PAR8_TRYBB
40 126.5 37.8 172 1 PRP2_RAT
41 126.5 37.8 243 1 TONE_PSEPU
42 126.5 37.8 514 1 MEFD_MOUSE
43 126.5 37.8 3119 1 HD_MOUSE
44 126 37.6 282 1 GDA3_WHEAT
45 126 37.6 3144 1 HD_HUMAN

P70315 mus musculu
O60885 homo sapien
P04726 triticum ae
P02831 mus musculu
O70324 mus musculu
P09791 trypanosoma
P10164 rattus norv
Q05613 pseudomonas
Q63943 mus musculu
P42859 mus musculu
P04723 triticum ae
P42858 homo sapien

ALIGNMENTS

RESULT 1

PERT_BORPA STANDARD; PRT; 922 AA.
AC P24328;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.70) (P.95).
GN PRN.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CN2591;
RX MEDLINE=91251771; PubMed=2041476;
RA Li L.J., Dougan G., Novotny P., Charles I.G.;
RT "P.70 pertactin, an outer-membrane protein from Bordetella
parapertussis: cloning, nucleotide sequence and surface expression in
Escherichia coli.";
RT Mol. Microbiol. 5:409-417(1991).
RL
CC -!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CONCENTRATIONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X54547; CAA38419.1;
CC EMBL; A26124; CAA01786.1;
CC EMBL; A19182; CAA01454.1;
CC PIR; S15204; S15204.
CC PIR; S14659; S14659.
CC InterPro; IPF004899; Pertactin_sup.
CC InterPro; IPF003992; pertactin.
CC InterPro; IPF003991; pertactin_vir.
CC Pfam; PF03212; Pertactin; 1.
CC PRINTS; PR01482; PERTACTIN.
CC PRINTS; PR01484; PERTACTINFAMLY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 922 P.95.
FT CHAIN 35 647 PERTACTIN (P.70).
FT PROPEP 648 922 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN

DR Pfam: PF00089; trypsin: 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYPSIN: 1.
 DR PROSITE; PS00240; TRYPSIN_DOM: 1.
 DR PROSITE; PS00134; TRYPSIN_HIS: 1.
 DR PROSITE; PS00135; TRYPSIN_SER: 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 431
 FT CHAIN 17 39
 FT CHAIN 40 2
 FT PROPEP 2 431
 FT DOMAIN 40 288
 FT DISULFID 22 152
 FT DISULFID 26 160
 FT DISULFID 71 87
 FT DISULFID 175 244
 FT DISULFID 207 223
 FT DISULFID 234 264
 FT CARBOHYD 19 19
 FT CARBOHYD 208 208
 FT ACT_SITE 86 86
 FT ACT_SITE 140 140
 FT ACT_SITE 238 238
 SQ SEQUENCE 431 AA; 46422 MW; 1C015A4E0BC0C68 CRC64;
 Query Match 44.9%; Score 150.5; DB 1; Length 431;
 Best Local Similarity 52.7%; Pred. No. 0.0036;
 Matches 29; Conservative 0; Mismatches 21; Indels 5; Gaps 2;

QY 2 AKAPPAPKPAQGPQP-GQPP-----QPPQPPQPPQPPQPPQPPQPPA 51
 DB 325 ASGPPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPH 379

RESULT 9
 ID GP1_CHLRE STANDARD; PRT; 555 AA.
 AC QPFP06; Q03927;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Vegetative cell wall protein gpi precursor (hydroxyproline-rich
 DE glycoprotein 1).
 GN Gpi.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21159092; PubMed=11258910;
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
 RA Goodenough U.W.;
 RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
 RT plant hydroxyproline-rich glycoproteins.";
 RL Biochemistry 40:2978-2987(2001).
 RN [2]
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=91017504; PubMed=1699225;
 RA Adair W.S., Apt K.E.;
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
 RT encoding cell wall hydroxyproline-rich glycoproteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
 CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
 CC layer.
 CC -1- SUBUNIT: Associates with GP2 and GP3.
 CC -1- PTM: N-glycosylated and O-glycosylated.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF309494; AAG45420.1; -;
 DR EMBL; M58496; AAA69706.1; ALT_SEQ.
 DR GlycoSuiteDB: Q9FP06; -;
 DR InterPro: IPR002965; P_Rich_extensin.
 DR InterPro: IPR003882; Pistil_extensin.
 DR PRINTS; PR01217; PRICHEXTENSIN.
 DR PRINTS; PR01218; PSTLEXTENSIN.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 555
 FT DOMAIN 40 339
 FT DOMAIN 259 279
 FT CARBOHYD 399 399
 FT CARBOHYD 455 455
 FT CARBOHYD 493 493
 SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
 Query Match 44.2%; Score 148; DB 1; Length 555;
 Best Local Similarity 54.0%; Pred. No. 0.006;
 Matches 27; Conservative 2; Mismatches 17; Indels 4; Gaps 2;

QY 4 APPAPKPAQGPQGPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 50
 DB 255 APPSPKP-PAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 303

RESULT 10
 ID Y066_NPVOP STANDARD; PRT; 875 AA.
 AC Q83949; Q65364; O10323;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 98.6 kDa protein (ORF71).
 OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=164623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97271300; PubMed=9126251;
 RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
 RA Rohrmann G.F.;
 RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
 RT polyhedrosis virus genome.";
 RL Virology 229:381-399(1997).
 RN [2]
 RP SEQUENCE OF 1-807 FROM N.A.
 RX MEDLINE=96201426; PubMed=8609478;
 RA Ahrens C.H., Rohrmann G.F.;
 RT "The DNA polymerase and helicase genes of a baculovirus of Orgyia
 RT pseudotsugata.";
 RL J. Gen. Virol. 77:825-837(1996).
 RN [3]
 RP SEQUENCE OF 806-875 FROM N.A.
 RX MEDLINE=95343549; PubMed=7618274;
 RA Ahrens C.H., Carlson C., Rohrmann G.F.;
 RT "Identification, sequence, and transcriptional analysis of lef-3, a
 RT gene essential for Orgyia pseudotsugata baculovirus DNA
 RT replication.";
 RL Virology 210:372-382(1995).
 CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV AND LDMNPV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

Qy 51 AG 52
|
Db 970 PG 971

Search completed: May 7, 2003, 16:48:11
Job time : 6.63203 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	335	100.0	922	2	S15204
2	284	84.8	911	2	A47675
3	209	62.4	910	2	A32560
4	168.5	50.3	430	2	JC2301
5	156	49.6	485	2	A33647
6	165.5	49.4	449	2	S16748
7	165	49.3	599	2	T10798
8	161	48.1	141	2	A34043
9	156	46.6	383	2	AE2295
10	152.5	45.5	487	2	S42442
11	152	45.4	409	2	A48232
12	150.5	44.9	431	2	S47538
13	150	44.8	439	2	S51939
14	150	44.8	1952	2	T48814
15	148.5	44.3	196	2	B48232
16	148	44.2	875	2	T10340
17	146.5	43.7	3164	1	WMBEH6
18	146	43.6	544	2	T17547
19	144.5	43.1	270	2	H83619
20	144.5	43.1	446	2	T07907
21	144	43.0	494	2	S96534
22	144	43.0	534	2	S21961
23	141.5	42.2	760	2	T06291
24	141.5	42.2	1137	2	A86335
25	141	42.1	440	2	I45681
26	141	42.1	929	2	C96623
27	140.5	41.9	1206	2	S24407
28	140.5	41.9	1468	2	T15151
29	140	41.8	839	2	F75518

A::Title: Presence in invertebrate genomes of sequences characterized by the repetition

A::Reference number: A90159; MUID:90147742; PMID:2105723

A::Accession: A34043

A::Molecule type: DNA

A::Residues: 1-141 <BAK>

A::Cross-references: GB:M32217

A::Accession: B34043

A::Molecule type: DNA

A::Residues: 59-136 <BA2>

A::Cross-references: GB:M32217

Query Match
Best Local Similarity 48.1%; Score 161; DB 2; Length 141;
Matches 27; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 5 PPAPKPAPQGPGQPQQPQQPPFPQQPQQPQQPEAPAPQPPAGR 53
||| || | || || || || || || || || || || ||
Db 11 PPPPPPppppppppppppppppppppppppppppppppppppr 59

RESULT 9
AE2295
hypothetical protein all3916 [imported] - Nostoc sp. (strain PCC 7120)
C::Species: Nostoc sp.
A::Note: Nostoc sp., strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C::Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
R::Accession: AE2295
R::Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara,
Nakazaki N.: Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A::Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A::Reference number: AB1807; MUID:21595285; PMID:11759840

A::Accession: AE2295

A::Status: preliminary

A::Molecule type: DNA

A::Residues: 1-383 <KUR>

A::Cross-references: GB:BA000019; PIDN:BAB75615.1; PID:gI7L33050; GSFDDB:GN00179

A::Experimental source: strain PCC 7120

C::Genetics:

A::Gene: all3916

Query Match
Best Local Similarity 46.6%; Score 156; DB 2; Length 383;
Matches 23; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

Qy 5 PPAPKPAPQGPGQPQQPQQPQQPQQPQQPQQPEAPAPQPP 50
||| || | || || || || || || || || || || ||
Db 337 PPDPppppppppppppppppppppppppppppppppppppr 382

RESULT 10
S42442
nuclear protein EBNA2 - human herpesvirus 4
C::Species: human herpesvirus 4, Epstein-Barr virus
C::Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C::Accession: S42442; S32988; S42447
R::Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
Proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986

A::Title: Nucleotide sequence of mRNAs encoding Epstein-Barr virus nuclear proteins: a p

A::Reference number: S42440; MUID:86259739; PMID:3460083

A::Accession: S42442

A::Molecule type: mRNA

A::Residues: 1-487 <SNM>

R::Farrell, P.J.

submitted to the EMBL Data Library, March 1988

A::Reference number: S32973

A::Accession: S32988

A::Molecule type: DNA

A::Residues: 1-487 <FAR>

A::Cross-references: EMBL:V01555; NID:G59074; PIDN:CAA24877.1; PID:gI632787

R:Dambaugh, T.; Hennessy, K.; Chammankit, L.; Kleif, E.
Proc. Natl. Acad. Sci. U.S.A. 81, 7632-7636, 1984

A::Title: U2 region of Epstein-Barr virus DNA may encode Epstein-Barr nuclear antigen 2.

RESULT 15
B48232

cysteine-rich extensin-like protein 2 precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 01-Dec-2000
C:Accession: B48232; PQ0474; S24616
R:Wu, H.; Zou, J.; May, B.; Gu, Q.; Cheung, A.Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 6829-6833, 1993
A:Title: A tobacco gene family for flower cell wall proteins with a proline-rich domain
A:Reference number: A48232; MUID:93342083; PMID:8341705
A:Accession: B48232
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-196 <WUA>
A:Cross-references: GB:I13440; NID:g310924; PIDN:AAA34060.1; PID:g310925
R:Ge S Goldman, M.H.; Pezzotti, M.; Seurinck, J.; Mariani, C.
Plant Cell 4, 1041-1051, 1992
A:Title: Developmental expression of tobacco pistil-specific genes encoding novel extensins
A:Reference number: PQ0474; MUID:93005740; PMID:1392807
A:Accession: PQ0474
A:Molecule type: mRNA
A:Residues: 'MAG', 1-105 <GOL>
A:Cross-references: EMBL:Z14014
A:Experimental source: stigma, style; strain Petit Havana SR1
C:Genetics:
A:Gene: CELP-2
C:Superfamily: glutelin
C:Keywords: cell wall; extracellular matrix; fertilization
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-196/Product: cysteine-rich extensin-like protein 2 #status experimental <MAT>

Query Match 44.3%; Score 148.5; DB 2; Length 196;
Best Local Similarity 50.9%; Pred. No. 0.0016;
Matches 27; Conservative 5; Mismatches 14; Indels 7; Gaps 2;

QY 6 PAPKPAQPQGPPGPQPQPQP---PQQPQPQPQPQR---QPEAPAPQPPA 51
| | | | | | | | | | | | | | | | | | | | | |
DB 63 PPRPRPRPRPCSPFPFPFRPCSPFPFPFRPCSPFPFPFRPSPPPPPPSPPPPA 115
| | | | | | | | | | | | | | | | | | | | | |

Search completed: May 7, 2003, 16:55:29
Job time : 12.8688 secs

GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-43

Query Match 41.2%; Score 138; DB 9; Length 538;
Best Local Similarity 51.7%; Pred. NO. 0.014;
Matches 30; Conservative 3; Mismatches 15; Indels 10; Gaps 4;
QY 2 AKAPPAPKPAQPGQPGQPO---PP-----QPQPPQPPQPPQPP-OROPEAPAPQPP 50
DB 178 AAPPPAAPPAVAPPAGRRAPPVAAAREPPLFP--PPQPPAPPPQPPQPPQPP 234

Search completed: May 7, 2003, 17:30:53
Job time : 18.3901 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 16:41:51 ; Search time 9.58433 Seconds
(without alignments)
178.054 Million cell updates/sec

Title: US-09-855-754B-19
Perfect score: 335
Sequence: 1 GAKAPPAPAPQPGQPGP.....QKQPEAPAPQPPAGRELSAA 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	335	100.0	922	4	US-08-460-269C-6
2	284	84.8	911	4	US-08-460-269C-4
3	209	62.4	910	4	US-08-460-269C-2
4	150	44.8	331	6	5202236-37
5	144	43.0	334	6	5202236-3
6	134.5	40.1	3119	1	US-08-246-982A-16
7	134.5	40.1	3119	1	US-08-453-265-16
8	128.5	38.4	106	4	US-09-314-268-134
9	127.5	38.1	1274	4	US-09-095-443-2
10	127	37.9	33	1	US-08-237-716-11
11	126.5	37.8	3118	2	US-08-457-273B-8
12	126	37.6	513	4	US-09-041-886-28
13	126	37.6	530	4	US-09-041-886-29
14	126	37.6	552	4	US-09-041-886-30
15	126	37.6	589	4	US-09-041-886-31
16	126	37.6	3144	1	US-08-246-982A-6
17	126	37.6	3144	1	US-08-453-265-6
18	126	37.6	3144	2	US-08-457-273B-42
19	126	37.6	3144	3	US-08-556-419-21
20	126	37.6	3144	4	US-09-041-886-15
21	125.5	37.5	60	1	US-08-534-342-12
22	125.5	37.5	60	1	US-08-675-140-12
23	125.5	37.5	478	3	US-08-155-888-2
24	124.5	37.2	174	4	US-08-818-112-143
25	124.5	37.2	174	4	US-08-818-111-138
26	124.5	37.2	174	4	US-09-056-556-143
27	124.5	37.2	174	4	US-09-072-596-138

28	124	37.0	24	4	US-08-750-624-11	Sequence 11, Appl
29	124	37.0	641	2	US-08-961-083-160	Sequence 160, App
30	122.5	36.6	485	2	US-08-749-391-2	Sequence 2, Appl1
31	122.5	36.6	485	4	US-09-390-200-2	Sequence 2, Appl1
32	122	36.4	330	1	US-08-642-255-32	Sequence 32, Appl1
33	122	36.4	408	1	US-07-609-716-65	Sequence 65, Appl
34	122	36.4	408	4	US-08-475-411A-65	Sequence 65, Appl
35	122	36.4	408	4	US-08-478-029A-65	Sequence 65, Appl
36	120	35.8	311	4	US-09-179-558-66	Sequence 66, Appl
37	120	35.8	351	2	US-08-929-417-2	Sequence 2, Appl1
38	119	35.5	1255	2	US-09-080-897-4	Sequence 4, Appl1
39	119	35.5	1255	3	US-08-899-595-1	Sequence 1, Appl1
40	119	35.5	1255	4	US-09-323-735-4	Sequence 4, Appl1
41	117.5	35.1	490	4	US-09-109-841-2	Sequence 2, Appl1
42	117	34.9	23	4	US-08-460-269C-8	Sequence 8, Appl1
43	117	34.9	357	1	US-07-609-716-66	Sequence 66, Appl
44	117	34.9	357	1	US-08-642-255-33	Sequence 33, Appl
45	117	34.9	357	4	US-08-475-411A-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-460-269C-6
; Sequence 6, Application US/08460269C
; Patent No. 6197548
GENERAL INFORMATION:
APPLICANT: CLARE, JEFFREY J.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
STREET: 2200 Clarendon Blvd., Suite 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6333
TELEFAX: (703) 243-6410
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 922 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 100.0%; Score 335; DB 4; Length 922;
Best Local Similarity 100.0%; Pred. No. 1.4e-21;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKAPPAPAPQPGQPGQPGP.....QKQPEAPAPQPPAGRELSAA 58
DB 564 GAKAPPAPAPQPGQPGQPGQPGP.....QKQPEAPAPQPPAGRELSAA 621

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


Db 38 PDGPDGPIPGGLNRLTSRNTSDPEGKCPSSLPPPPPPPTTPPECGEGHPPP 97
QY 47 POPPAGE 54
Db 98 PPPNGHD 105

RESULT 9
US-09-095-443-2
; Sequence 2, Application US/09095443
; Patent No. 6342593
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, Gregory
; APPLICANT: PELES, Elor
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
; OF ALP RELATED DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON 5 LYON
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: California
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FASTSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,443
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/049,477
; FILING DATE: June 12, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 235/055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-095-443-2

Query Match 38.1%; Score 127.5; DB 4; Length 1274;
Best Local Similarity 43.5%; Pred. No. 0.00074;
Matches 37; Conservative 1; Mismatches 12; Indels 35; Gaps 7;

QY 1 GAKAP-----PAPQGPQP-----GPPQPPQPPQPPQ--PP 35
Db 565 GAKQPIPAQHFFSGITGFPAPRIGPQHPQHPHSQAQFGPQPQLPLQHPHLP 624

QY 36 Q-----PPQR-QPEAPAP-----QPP 50
Db 625 QAPGLLPQSPYPAPQGVLGQPP 649

RESULT 10
US-08-237-716-11
; Sequence 11, Application US/08237716
; Patent No. 5589384

GENERAL INFORMATION:
APPLICANT: LIPSCOMBE, Martin J
APPLICANT: CHARLES, Ian G
APPLICANT: FAIRWEATHER, Neil F
TITLE OF INVENTION: FUSION PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye, P.C.
STREET: 1100 No. 5589384th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/237,716
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,003
FILING DATE: 11-JUN-1992
APPLICATION NUMBER: GB 9112553.4
FILING DATE: 11-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mary J
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-157
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1-2
OTHER INFORMATION: /label= LTB sequence
FEATURE:
NAME/KEY: Region
LOCATION: 4..7
OTHER INFORMATION: /label= hinge region
FEATURE:
NAME/KEY: Region
LOCATION: 10..31
OTHER INFORMATION: /label= p69 BB05 epitope of Bordetella pertussis sequence
US-08-237-716-11

Query Match 37.9%; Score 127; DB 1; Length 33;
Best Local Similarity 57.5%; Pred. No. 3.3e-05;
Matches 23; Conservative 2; Mismatches 3; Indels 12; Gaps 1;

QY 10 PAPQGPQGPQGPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 49
Db 5 PGPEIAPQGPQGPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 32

RESULT 11
US-08-457-273B-8
; Sequence 8, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biayang
; APPLICANT: Nasir, Jamal

